

#OvarianCancer: Using Social Media to Facilitate Health Communication

Tara Coffin

A dissertation

submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Deborah Bowen, Chair

Malia Fullerton

Hilary Karasz

Program Authorized to Offer Degree:

Public Health Genetics

©Copyright 2019

Tara Coffin

University of Washington

Abstract

OvarianCancer: Using Social Media to Facilitate Health Communication

Tara Coffin

Chair of the Supervisory Committee:

Deborah Bowen

Bioethics & Humanities

Ovarian cancer is the fifth leading cause of cancer death among women. Women with a personal or family history of breast and ovarian cancer are encouraged to seek genetic testing, but fewer than 15 to 30% of eligible patients are ever offered genetic testing. Social media may be well positioned to address this area of need. This research seeks to improve our understanding of social media as a platform for public health research and communication, in the context of familial ovarian cancer susceptibility. In Aim 1, I explored how ovarian cancer and ovarian cancer risk is discussed over Twitter, collecting and analyzing tweets to identify structural features of a tweet and describe how social media users publicly exchange and engage with ovarian cancer information over Twitter. In Aim 2, I conducted a series of contextual interviews to describe the online health information-seeking behaviors of internet users at risk for ovarian cancer. Finally, in Aim 3, I evaluated innovative methods for recruiting research participants for an ovarian cancer genetic testing study, using paid targeted advertisements and unpaid posts over Facebook. Findings from this research illustrate what structural components of a social media

post may help improve message reach, describe how internet users with a risk of ovarian cancer assess online health information, and demonstrate that Facebook is an effective tool for reaching women over the age of 30 with an increased risk of ovarian cancer. These findings offer insight into how public health professionals can leverage social media for the purposes of health communication, and demonstrate that with the help of patient advocates, online communication tools like social media may improve representation in research and health service provision, by widening messaging reach.

Table of Contents

List of Tables	7
List of Figures	8
Introduction	9
Summary of aims	11
Background	12
Innovation	21
Overview of Research	23
Aim 1	23
Aim 2	23
Aim 3	24
Aim 1	25
Introduction	25
Methods	31
Results	37
Discussion	44
Limitations	48
Aim 2	51
Introduction	51
Methods	56
Results	61
Discussion	74
Limitations	81
Aim 3	82

Introduction	82
Methods	88
Results	97
Discussion	103
Limitations	108
Conclusion	112
Appendix	120
References	148

List of Tables

Table 1. Summary of total Tweets and user accounts	124
Table 2. Summary of the structural components observed across human- and social bot-generated Tweets.	125
Table 3: Top ten most retweets from human operated twitter accounts.	126
Table 4: Summary of the thematic analysis of human-generated Tweets and social bot-generated Tweets	127
Table 5: Adapted eHEALS questionnaire used in part one of interview	129
Table 6: Interview task scenarios	130
Table 7: Follow up interview questions	131
Table 8: Mean scores for the adapted Individual eHEALS items and other section one questions	132
Table 9. Facebook analytic variables defined (also will be defined in text)	138
Table 10. Description of Facebook paid campaign content and audience	140
Table 11. Summary of results for all paid campaigns	144
Table 12. Summary of enrollment and randomization data.	147

List of Figures

Figure 1: Example Tweet, including an external link, and text content	120
Figure 2: Study flow diagram detailing the flow of Aim 1	121
Figure 3. Sample of Botometer report.	122
Figure 4. Plot of the eligible human and social bot generated Tweet count over the period from May 10 to May 23, 2017	123
Figure 5: Consort flow diagram of study	133
Figure 6: Example of an interactive online risk assessment quiz participants visited.	134
Figure 7: Illustration of the online communication system utilized by the MAGENTA study.	135
Figure 8: Example of paid Facebook advertisement used by the MAGENTA study.	136
Figure 9: Example of unpaid Facebook post published to the MAGENTA study Facebook page.	137
Figure 10. Timeline of enrollment trends and recruitment events.	143

Introduction

There is a well-established link between pathogenic variants (and other genetic alterations) in the *BRCA1* and *BRCA2*, as well as other cancer-linked susceptibility genes, and an increased risk of ovarian cancer (Howlander et al., 2017; Kuchenbaecker et al., 2017; Campeau et al., 2008; Walsh et al., 2006). Many pathogenic variants, including *BRCA1* and *BRCA2*, that place women at a higher risk for ovarian cancer, also raise an individual's risk of breast cancer (Howlander et al., 2017; Kuchenbaecker et al., 2017). As a result, women with a personal or family history of breast and ovarian cancer are encouraged to seek genetic testing (Moyer, 2014). Those who receive genetic testing are better able to make informed decisions about prevention, early detection, and treatment (National Cancer Institute, 2018). Despite these recommendations however, there are gaps in the availability of genetic testing. Currently, fewer than half of eligible women diagnosed with breast cancer receive genetic testing (Moyer, 2014). There are a variety of hurdles that prohibit access to, and uptake of, genetic testing services, including financial and structural barriers (Anderson et al., 2012; Trivers et al., 2011; Acheson et al., 2005). Eligible individuals may also not have the opportunity to decide if genetic testing is right for them. Too often, healthcare professionals fail to refer at-risk-patients, and healthcare clinics aren't always equipped to offer specialized services (Anderson et al., 2012; Brandt et al., 2008). As a result, only 15 to 30% of eligible patients, defined as individuals with ovaries who have a personal or family history of breast or ovarian cancer, are ever offered genetic testing (George et al., 2016; Febbraro et al., 2015; Ricci et al., 2015; Slade et al., 2015; Norquist et al., 2013). Testing uptake is even lower for low income and minority populations (Saulsberry & Terry, 2013).

Demand for genetic testing services is slated to increase with current genetic testing recommendations for individuals with a personal or family history of breast and ovarian cancer. As service availability lags, public health professionals are now facing an urgent need to explore alternatives to increase awareness of genetic testing, empowering at risk individuals to ask about services even when their doctor does not (Schwartz et al., 2014; Bradbury et al, 2010). Current research suggests that social media may be a good channel to improve public awareness of genetic testing services, as the percentage of individuals with social media access continues to rise (Duggan et al., 2015; Bonevski et al. 2014). Online social media addresses many of the hurdles that keep specialized health services, like genetic testing, out of the public eye, including clinic limitations and geographic barriers (Martinez et al., 2014). At the same time, the use of social media as a health communication tool may pose some risks.

Today, the majority of adult internet users go online with their health questions, turning to the internet before contacting their health care provider (Huerta et al., 2016; Volkman et al., 2014; Duggan et al., 2015; Morgan et al., 2014; Kuehn, 2013). This is also true among individuals seeking information about cancer (Huerta et al., 2016; Volkman et al., 2014). Online social media is well positioned to leverage this trend, facilitating targeted information dissemination over a platform where users can rapidly seek, generate, engage with, and share health information online. The term social media refers to Internet-based tools that allow users to communicate with others, sharing and engaging with information between peers (Ventola, 2014). Social media tools include a variety of online social networking platforms, blogs, media sharing sites, virtual reality, and gaming environments. These platforms offer public health researchers unique, and often cost-effective, opportunities for research recruitment, particularly for highly stigmatized topics or connecting with populations that are hard to reach, like those impacted by a

rare disease (Arigo et al., 2018). But there are lingering concerns about the quality and reliability of online health information, despite potential benefits associated with exchanging health information online and over social media (Friedman et al., 2011; Kukreja et al., 2011; Orizio et al., 2010; Tian, 2010). Without an adequate understanding of social media as a health communication tool, public health professionals are often left uncertain about how best to harness social media and reach targeted populations (Moorhead et al., 2013; Correa et al., 2010).

There is a growing need to explore different approaches to social media for the purposes of health communication, particularly in the context of hereditary ovarian cancer risk and genetic testing (Moorhead et al., 2013). By better understanding how users interact with social media for health information seeking purposes, health professionals will be better equipped to utilize this innovative communication tool (Heron & Smyth, 2010; Patrick et al., 2005; Intille et al., 2003). This research seeks to improve our understanding of social media as a platform for public health research and communication, in the context of familial ovarian cancer susceptibility, by:

- **Aim 1:** Exploring how ovarian cancer is discussed over Twitter, collecting and analyzing tweets to identify structural features of a tweet and describe how social media users publicly exchange and engage with ovarian cancer information over Twitter.
- **Aim 2:** Describing the online health information-seeking behaviors (HISBs) of internet users with a known genetic predisposition for ovarian cancer.
- **Aim 3:** Evaluating innovative methods for recruiting research participants for an ovarian cancer genetic testing study, over Facebook, including the use of paid targeted advertising and unpaid posts.

Background

Women with a personal or family history of breast or ovarian cancer are encouraged to seek genetic testing.

Current recommendations urge women with a personal or family history of breast or ovarian cancer to seek genetic testing (Moyer, 2014). With an established genetic association between cancer-linked susceptibility genes and an increased risk of ovarian cancer (Howlander et al., 2017; Kuchenbaecker et al., 2017; Campeau et al., 2008; Walsh et al., 2006), the hope is that genetic testing may improve outcomes for at risk individuals via risk reduction options. For example, an individual may opt to reduce their risk of ovarian cancer by undergoing an oophorectomy (Finch et al., 2014; Finch et al., 2011). An oophorectomy may not be an appropriate option for younger women interested in maintaining their reproductive health, or for those disinterested in the prospect of early menopause. Genetic testing for ovarian cancer provides low clinical utility for these individuals, but they may still benefit from receiving genetic testing, given the potential psychological, social, and practical benefits (Bunnik et al., 2015). For instance, those who learn they have not inherited a known pathogenic variant segregating in the family may experience relief (National Cancer Institute, 2018), and positive results may inform health decision making and encourage testing of other at risk family members (Bunnik et al., 2014; Kopits et al., 2011). Given the potential clinical and personal utility, current guidelines recommend genetic testing for at-risk individuals and the demand for genetic testing is on the rise (Schwartz et al., 2014; Bradbury et al, 2010).

Genetic testing uptake is low among those at risk for ovarian cancer.

An estimated 15% of all ovarian cancer patients have a germline pathogenic variant, often referred to as a mutation, in the *BRCA1* and *BRCA2* genes, with implications for their

personal health management, as well as that of their family (George et al., 2016). Despite this statistic and current recommendations in favor of genetic counseling, only 15 to 30% of eligible patients are ever offered testing (George et al., 2016; Febbraro, et al., 2015; Ricci et al., 2015; Slade et al., 2015; Norquist et al., 2013). Low uptake of genetic testing services may be due to any number of factors, including financial barriers, structural barriers, or lack of being told that genetic testing and counseling services are even available (Febbraro, et al., 2015; Petzel et al., 2014; Anderson et al., 2012; Trivers et al., 2011). If an at-risk individual never learns that genetic testing may be a useful healthcare tool, they are unlikely to seek out, and benefit from, this valuable health service. Public health professionals are facing an urgent need to explore alternative approaches to health communication as genetic testing availability and uptake continues to lag (Schwartz et al., 2014; Bradbury et al, 2010).

Most adult internet users go online for health information.

Social media are communication tools that facilitate the flow of information between peers, health professionals, and the public. The term social media encompasses online platforms such as Pinterest, Instagram, Twitter, YouTube, SnapChat, and Facebook, with new platforms rapidly emerging. These platforms interact with online search engines and improve reach through peer-to-peer information exchange, an interplay which defines Web 2.0 and has transformed health communication patterns (Tennant et al., 2015; Gibbons et al., 2011). Prior to the wide adoption of social media (i.e. Web 1.0), much of the health information available online closely resembled vertically transmitted information, with a focus on top-down one-way communication rooted in objective facts from a source of authority (Newby et al., 2015; Aghaei et al., 2012). In contrast, social media facilitates a forum for discussion, where information is also shared horizontally between peers (Newby et al., 2015; Tennant et al., 2015; Keller et al.

2014). Anyone with an internet connection is invited to create, exchange, and consume user-generated content, transforming passive consumers into content producers with increasing frequency (Keim & Noji 2011; Arigo et al., 2018).

Today, over 72% of adult internet users go online for health information (Volkman et al., 2014; Fox & Duggan, 2013). This trend of seeking and consuming health information over social media is slated to increase along with growing mobile device ownership (PEW Research, 2018C). Social media users feel that the internet is a valuable source for health information, and often query their health concerns over social networking sites inhabited by people with shared health experiences (Duggan et al., 2015; Hong et al., 2012). Many of these individuals use the internet to address their health concerns and even self-diagnose before contacting their health care provider (Morgan et al., 2014; Kuehn, 2013; Fox, 2011; Siliquini et al., 2011), improving access by sharing and seeking health information online and over social media (Habibi et al., 2012; Macnab et al., 2012; Rojjanasrirat et al., 2012; Greene et al., 2011). Users can also overcome barriers that may have prevented access to necessary health information or services like genetic testing, including geographic isolation, transportation hurdles, work hours, lack of childcare, and lack of appropriate healthcare services (Griffiths et al., 2012; Sadler et al., 2010; Bowen et al., 2004; Doherty, 2004; PENCHANSKY & THOMAS, 1981). As inequities in Internet availability and accessibility continue to diminish, more and more adult internet users are able to access social media platforms, with little difference in the education level, income, or developed environment observed across adult social media users (PEW Research, 2018C; Greenwood et al., 2016). Researchers have some understanding of how individuals with cancer interact with online resources and social media for information seeking purposes, but little is known about those with

a family history of ovarian cancer or a genetic predisposition to ovarian cancer (Lambert et al., 2009; Roach et al., 2009; Rutten et al., 2005).

Social media facilitates information seeking and scanning.

Online knowledge acquisition like this is traditionally divided into two distinct categories—information seeking and information scanning (Niederdeppe et al., 2007). Social media facilitates both, shifting the information seeking terrain into a more passive territory by blurring the line between health information seeking and health information scanning (Khoo, 2014; Setoyama et al., 2011). Health information seeking occurs when an individual actively reaches out for information in some fashion and includes any goal-oriented use of a resource (Niederdeppe et al., 2007). Information seeking encompasses instances where an individual looks for a specific type of health information, as opposed to incidentally finding it when casually surfing the internet. Health information scanning, on the other hand, is defined as the ascertainment of health information through passive exposure during routine use of a resource or conversation (Niederdeppe, et al., 2007). Compared to information seeking, health information scanning is less goal-oriented. Health information scanning includes any instance where a user comes across health information through routine use of their social media account. The benefits of health information scanning over social media appears to mirror the effects of health information seeking. Health information scanning may also improve the reach of a research related message over social media, particularly through the use of paid social media advertisements.

Social media poses many challenges and opportunities as a public health communication tool.

Peer-to-peer information exchanges result in interactions that fail to account for quality of information, resulting in a mixture of evidence-based and practice-based knowledge, the latter which is fueled by anecdotal evidence (Moorehead et al., 2016). Studies exploring the quality of online cancer information draw attention to the lack of accuracy and quality assurance across the majority of online health information sources (Banki et al., 2016; Sobota & Ozakinci, 2015; Kukreja et al., 2011; Friedman et al., 2011). Inaccurate information may travel quickly over social media, leaving users with the responsibility to verify sources (Craigie et al., 2002). But users may be disinclined to question information they receive from a trusted or familiar source, like a friend or loved one (Steelman et al., 2015). This problem is compounded further by what Quattrociocchi et al. (2016) referred to as the “Echo Chambers” on social media. This concept is based on the idea that social media users tend to promote their favorite narratives. This enables users to process information from the Internet selectively, in an unbalanced, biased fashion, forming a self-serving image of their own health (Sassenberg et al., 2016). In other words, an individual’s social network has the potential to impact the type of online information they have access to, in regards to ovarian cancer and genetic testing.

The exchange of inaccurate health information has the potential to do harm—misleading decision making, encouraging harmful health practices, and may negatively impact the physician-patient relationship (Ventola, 2014; Lau, et al., 2012; Murray, et al., 2003). Several studies have sought to understand how misinformation over social media may impact health outcomes. To date, few studies have found evidence of measurable harm (Noilke et al., 2015; Syed-Abdul et al., 2013; Singh et al., 2012), with research speaking favorably of the public’s

ability to discern the quality of online information (Cole et al., 2016). However, these studies draw on small sample sizes and fail to adequately address how online health information may impact public health communication efforts. For example, opponents of certain health services, like vaccination, use social media as a mechanism to organize opposition and extend the reach of their message (Bean, 2011; Larson et al., 2011; Cooper et al., 2008). With the help of Facebook and Twitter, anti-vaccination groups are able to connect directly with individuals seeking information about vaccines online, using their vantage point to evoke concerns and continue the spread of misinformation through content published to social media platforms (Mitra et al., 2016). In the case of vaccines, inaccurate information over social media may lower rates of vaccine compliance and contribute to the resurgence of vaccine preventable diseases (Mast et al., 2015). Such outcomes pose a disadvantage to providers and patients alike. Given these observations, it is important to further consider how the quality of online health information, including content shared over social media, may impact outcomes and access to genetic testing.

Internet users still benefit from interacting with, and sharing, health information virtually, regardless of quality or accuracy (Hong et al., 2012; Setoyama et al., 2011; Høybye et al., 2005). The more social media users interact with health-related content over Facebook or Twitter, the more likely it is that other users will see it when engaging in information seeking and scanning behaviors. As more users have access, more users socially engage with content, and essentially promote it, through shares and “likes,” tailoring health information along the way to better fit user needs (McLaughlin et al., 2012; Lee et al., 2010; Shigaki et al., 2008), and health information seekers are able to participate in medical decision making and are more motivated to make behavior changes (Ventola, 2014). The simple act of exchanging health information online may also mitigate feelings of isolation, address the stress associated with a health concern,

facilitate a forum to share experiences, and shift user attitudes in regard to their capacity to control their own health (Greene et al., 2011; Setoyama et al., 2011; Freyne, et al., 2010; Høybye et al., 2005; Eysenbach, 2003).

Social media is well positioned to improve ovarian cancer research and communication efforts, but more research is needed.

Public health researchers can also benefit from exploring the virtual space of social media, particularly related to study recruitment. Current research suggests that social media may offer affordable solutions to improving research recruitment efforts, enabling researchers to target specific populations and adapt recruitment efforts quickly to meet study requirements (Pedersen & Kurz, 2016; Fenner et al., 2012). Social media also facilitates respondent driven sampling, or snowball sampling, and repeat exposure (Bauermeister et al., 2012). Respondent driven sampling utilizes the peer network inherent to an online social media platform, encouraging peers to interact with, and share, information related to a public health research effort or associated topic (Pedersen and Kurz, 2016). It is this peer network that makes social media particularly well suited for health communication and research efforts in the arena of hereditary cancer. In other words, recruitment efforts can be improved when peer networks of ideal research participants share information about study participation opportunities over social media.

To date, various research efforts have explored how technology may improve access to genetic testing. For example, a number of studies have found that phone-based genetic counseling services can improve genetic testing uptake (Schwartz et al., 2014; Pal et al., 2010). Other research efforts have focused on the role handheld mobile devices play in the provision of health services in general, using eHealth and mHealth approaches to improve access to health

services among underserved rural communities (Silva et al., 2015). Current trends in handheld mobile device ownership and online health information seeking suggest that social media may be a suitable communication tool to address barriers to the uptake of ovarian cancer genetic testing. But the use of social media to build public awareness of genetic testing is largely unexplored.

The adoption of social media for health communication purposes has been slow, despite potential benefits. Public health professionals remain concerned about quality and reliability of online health information (Banki et al., 2016; Sobota & Ozakinci, 2015; Kukreja et al., 2011; Friedman et al., 2011), and have reservations about using social media in the research setting, worried that this modern communication tool may fail to recruit a representative sample (Pedersen & Kurz, 2016) or may not be appropriate for recruiting older adults, including those over the age of 30 (Whitaker et al., 2017; Kapp et al., 2013). Further, with few resources speaking to best practice, public health professionals are uncertain about how valuable this tool is for reaching targeted populations, let alone how best to harness it (Moorhead et al., 2013; Correa et al., 2010). Given the increase in mobile device ownership, consumer benefit, and the subsequent move towards eHealth and mHealth practices, there is a growing need to explore approaches to, and the relative effectiveness of, different applications of social media for health communication (Moorhead et al., 2013).

More research is needed to better understand how public health professionals can use social media for health communication purposes.

More research is needed to enable public health professionals to leverage online social media and improve access to ovarian cancer information and genetic testing services (Moorhead et al., 2013; Correa et al., 2010). This study describes and evaluates effective approaches to using social media to communicate information about ovarian cancer by exploring how the public uses

popular social media platforms to seek and exchange information about ovarian cancer and assessing social media as a recruitment tool targeting women at risk of ovarian cancer. Much of this proposed research revolves around a selection of specific social media platforms, specifically Facebook and Twitter. Lafferty & Manca (2018) suggests that platform specific research like this may not transfer to other social media platforms, a challenge that is particularly problematic given the fast-evolving terrain of social media platforms, with new platforms emerging almost daily. This research keeps this evolution in mind, identifying what components of a social media post improve the message reach and uptake, generating findings that are not tied to specific social media platforms. These structural components can be used across other social media platforms for health communication and research recruitment purposes.

Innovation

Using social media as an innovative health communication tool. Social media is an innovative communication tool for sharing health information. As a result of established peer-to-peer networks, social media is primed for respondent driven sampling. Some studies have sought to better understand how this tool functions in the community setting, but little research has explored how social media may be used to improve communication efforts and address the deficit in genetic testing uptake among women at risk of ovarian cancer. This research addressed this deficit, adding to what is known about social media as a health communication tool in the context of ovarian cancer genetic testing.

Using social media as an innovative recruitment tool. Social media is also an innovative recruitment tool, enabling public health researchers to target specific populations for research recruitment. Social media has been well explored as a recruitment tool to reach certain populations but has not been adequately tested as a means of reaching those who may be at higher risk of ovarian cancer or individuals impacted by certain types of hereditary cancer predisposition who may benefit from participating in genetic testing.

Performing contextual data collection. Social media presents researchers with a unique opportunity to observe real-time data in context. Where many traditional research methods rely on participant recall, by collecting data directly from social media platforms this research collected data that is not subject to recall bias or the Hawthorne effect (McCarney et al., 2007). Twitter data, which is publicly available and searchable, is particularly unique and useful for researchers. Given the public nature of tweets, Twitter users do not have an expectation of privacy and researchers can openly observe the resulting online content, enabling the quick collection of a large amount of publicly available data for analysis.

Contextual interviews as an innovative way to learn about information seeking

behaviors. Contextual interviews require participants to respond to a prompt, or in this case an information seeking task, in a context that they may naturally respond to the scenario in reality. The approach used in this research is a unique method for studying online information seeking behaviors. Other studies exploring online information seeking behaviors rely primarily on participant recall and are subject to recall bias. By conducting contextual, task driven interviews I had a unique opportunity to observe and describe information seeking behaviors in action of individuals with a confirmed genetic predisposition to ovarian cancer. Not only did this offer insight into how at-risk individuals seek information about ovarian cancer, but it also shed light on how contextual interviews function as a tool for exploring online health information seeking behaviors.

Overview of Research

Aim 1: *Exploring how ovarian cancer is discussed over Twitter, collecting and analyzing tweets to identify structural features of a tweet and describe how social media users publicly exchange and engage with ovarian cancer information over Twitter.*

Aim 1 targets explores how the public currently uses social media, specifically Twitter, to disseminate information about ovarian cancer. By exploring how ovarian cancer is mentioned and discussed over Twitter, I developed a method for identifying Twitter social bots and identified what structural components of a social media message impact its transmission, shaping best practice recommendations for building a social media communication plan. These recommendations include notes about the use of imagery, hashtags, tone, and external links, among other structural features that impact the reach of a social media post (Chiang et al., 2016; Comm, 2009).

Aim 2: *Describing the online health information-seeking behaviors (HISBs) of internet users with a known genetic predisposition for ovarian cancer.*

Aim 2 expands upon this investigation, using the Comprehensive Model of Information Seeking (CMIS) to directly observed the online information seeking behaviors of individuals who have been told by a health care provider that they have an increased risk of developing ovarian cancer. The CMIS model has been widely used to explain the information seeking behaviors of cancer information seekers, drawing on concepts and findings from a synthesis of the Health Belief Model (Rosenstock, 1974), Uses and Gratification Theory (Rubin, 1986), and the Media Exposure and Appraisal Model (Johnson et al., 1993A & 1993B) (for an extensive discussion, see Johnson, 1997). Participants responded to a series of prompts over the course of the interview. By observing their online behaviors, I identified what characteristics of an online

source elicit user trust and learned where an individual with a risk of ovarian cancer may look for online information about ovarian cancer. Like Aim 1, these observations may inform best practice recommendations, with a focus on components that make an online source more or less trustworthy to the public and identify where individuals are most likely to look for ovarian cancer information over the Internet.

Aim 3: *Evaluating innovative methods for recruiting research participants for an ovarian cancer genetic testing study, over Facebook, including the use of paid targeted advertising and unpaid posts.*

Findings from Aim 1 and Aim 2 informed online health communication efforts in the context of ovarian cancer and ovarian cancer genetic testing. These findings also informed Aim 3, which focused on evaluating the efficacy of social media as a recruitment tool for an ovarian cancer genetic testing study. For this aim, information was pushed out over social media, through targeted paid Facebook advertisements and unpaid Facebook posts. The method to achieve this aim utilized Facebook, but the goal was to piece together successful components of online health communication, independent of a specific social media platform.

AIM 1

Introduction

Background

Ovarian cancer is the fifth leading cause of cancer death among women. The American Cancer Society estimates that more than 22,000 women will receive a new diagnosis of ovarian cancer in 2018 and over 14,000 will die from ovarian cancer, accounting for more deaths than any other cancer of the female reproductive system (Siegel et al., 2018). An estimated 15% of all ovarian cancer patients possess a germline pathogenic variant, or a hereditary gene change, in the *BRCA1* and *BRCA2* genes. Unfortunately, only 15 to 30% of eligible patients are ever offered testing despite the fact that genetic testing may improve outcomes for women at risk for ovarian cancer by describing risk and informing health decision making (National Cancer Institute, 2018; George, et al., 2016; Febbraro, et al., 2015; Ricci et al., 2015; Slade et al., 2015; Norquist et al., 2013). Those who may benefit from genetic testing are left in the dark without effective communication and genetic testing uptake remains low (George, et al., 2016; Petzel et al., 2014). With these trends in mind, public health professionals are increasingly interested in improving communication efforts surrounding ovarian cancer and associated genetic testing (Schwartz et al., 2014; Bradbury et al, 2010). Social media platforms may be well positioned to address this deficit as the public increasingly engages with health information online.

Social media are broadly defined as a group of online applications, or social platforms, that enable users to create, share, and engage with content online. Around 65% of American adults report ownership of at least one active social media account (Perrin, 2015). Research indicates that these ubiquitous online tools impact how the public engages with cancer-related health information (Shen et al., 2015; Basch et al., 2004; Monnier et al., 2002). For example, the

move toward online social networking has shifted the vertical transmission of health information, from physician to patient, towards a horizontal exchange of information between peers (Newby et al, 2015; Ancker et al., 2009). Today, an increasing percentage of patients with cancer, cancer survivors, and caregivers use the internet to seek and engage with cancer-related health information (Im et al., 2018; Chou et al., 2011). This trend includes ovarian cancer information seekers, who use web-based message boards and social media to address their health information seeking needs (Ancker et al., 2009; St-Pierre et al., 2018). Facing Our Risk Together, for example, provides online discussion boards for peer-to-peer information exchange with a focus on ovarian cancer (www.FacingOurRisk.org; Ancker et al., 2009). Peer-to-peer information exchange like this empowers the public to play a more active role in sharing health information, improving accessibility by circumventing traditional gatekeeping of health professionals and shaping online health communication and content (Ventola, 2014; MacMillan, 2012).

This peer-to-peer information exchange prompts concern about the quality of information, particularly given how quickly misinformation can travel online (Banki et al., 2016; Sobota & Ozakinci, 2015; Friedman et al., 2011; Kukreja et al., 2011; Adams, 2010A & 2010B). Misinformation may increase worry, negatively impact decision making, or leave the individual without critical information about preventative measures like genetic testing (Miller & Bell, 2012; Zulman, et al., 2011; Arora et al., 2008; Hesse et al., 2005; Eysenbach et al., 2002). These concerns warrant investigation. But increased access to health information online also has the potential to assist public health research endeavors and communication efforts (Merchant et al., 2011; Farmer et al., 2009), inform public health investigations (Yin et al., 2015), and address public health issues (Paul & Dredze, 2011; Scanfeld et al., 2010). As social media use continues

to increase, so does the potential value to public health communication, including the use of social media as a health communication tool in the context of ovarian cancer.

Twitter, in particular, provides a unique opportunity for learning how the public shares and engages with health information. Twitter (www.Twitter.com) is a popular online microblogging website, with over 67 million active monthly users from the US (Twitter, 2018a). Users post publicly available comments, or Tweets, containing short snippets of content, with structural components like text, images, and links to other websites, interacting and sharing other user generated content. Prior research suggests that the use of certain structural components, such as imagery or trending hashtags may improve the reach of health information, enabling lay audiences to more easily identify quality health information online (Chiang et al., 2016; Comm, 2009). These structural components include, but are not limited to:

1. Hashtags: a keyword or phrase that is preceded by a “#”
2. Images or media: a picture, illustration, or video embedded into a post
3. Call to action: a direct prompt directing the users to perform an action which usually involves clicking a link or sharing content
4. @mention: tagging another user in the Tweet or social media post by typing an @ sign, followed by their user name
5. External link/URL: a URL to another webpage for social media users to click on

Tweets are limited to 140-characters and can be read without owning a Twitter account (Lafferty & Manca, 2018). Other users can engage with this content. Social media engagement includes any time a user responds to or comments on a post, shares or retweets it, meaning they post it to their own social media page, or like it (Twitter, 2018b). Users may like a post via a built-in mechanism on the social media platform that allows users to show their agreement or

support of a social media post with a click of a button. Each type of engagement helps facilitate rapid dissemination (Twitter, 2018b). Twitter content is publicly available and searchable, a quality that is highlighted by Twitter's privacy policy, which states, 'What you say on Twitter may be viewed all around the world instantly' (Twitter, 2014a & 2014b). In other words, researchers can openly observe user-generated content over Twitter (Burton et al., 2012).

Data generated from Tweets bear similarities to data produced through surveys or structured observations. However, there are some novel advantages to data generated over social media when compared to survey data or data gathered from structured observations. For example, while a survey asks participants to retrospectively recall behaviors, social media facilitates large scale observation of participant interaction and unprompted information sharing in real time. Unlike traditional observational and ethnographic research, with Twitter researchers can preserve and review interactions as they occurred, rather than relying on their perceptions of what transpired (McCormick et al., 2015). Instead of relying on field notes, which may be subject to observer bias, Twitter data preserves social interactions and enables researchers to review those interactions directly.

Twitter is not without its faults as a data source. Like other social media platforms, Twitter is subject to third-party manipulation through the presence of social robots, or social bots. Social bots are algorithmically-generated and largely automated social media profiles built to look like actual human users, with profile pictures and location details. Social bots rely on computer generated algorithms to generate Tweets, influencing discussion and promoting specific ideas or products (Ferrara et al., 2016). Many social bots differ from their human counterparts in the frequency of Tweets and indiscriminately reference content, but more sophisticated bots produce content that appears to be human-generated and can interact with

legitimate human accounts. Social bots are credited with influencing politics (Bessi & Ferrara, 2016), stock market manipulation (Ferrara, 2015), promotional content (Varol et al., 2017a), and the spread of misinformation (Varol et al., 2017b). Social bots are responsible for two-thirds of the links shared on Twitter, so it is important to distinguish social bot Twitter accounts from human Twitter accounts to explore differences in online information sharing behaviors between these two account types (Wojcik et al., 2018; Allem et al., 2017).

Another disadvantage to using Twitter as a data source is that Twitter users are not representative of the general population (Bruns & Burgess, 2012; Bruns & Stieglitz, 2013; Giglietto et al., 2012; Heavilin et al., 2011; King et al., 2013; Ovadia, 2009). The typical Twitter user is a young, affluent, college-educated, non-white male (Duggan et al., 2015). On average, Twitter users are between the ages of 18 and 29 (37%) and 30 to 49 (25%). Most Twitter users have a college degree, come from urban settings, and over half of user's report making over \$50,000 per year (Duggan, 2015). Nevertheless, there is much to gain by examining how ovarian cancer is discussed over Twitter.

Aims of the study

The purpose of this study is to develop an understanding of how ovarian cancer is referenced over Twitter, with the goal of identifying the ways this tool may be leveraged by public health professionals for communication purposes. This study collected and analyzed publicly available Tweets containing the term "ovarian cancer" over a two-week period of time, to (1) develop an understanding for how Twitter accounts reference ovarian cancer and (2) identify what structural components of a social media message impact its transmission. By analyzing Tweets mentioning ovarian cancer, researchers can identify common structural components of social media posts, and better understand how these types of components impact

social media communication efforts (Smith & Brenner, 2012; Comm, 2009). This study will also explore if the use of these components differ across human-generated content and social bot-generated content. The concept of a hashtag is native to Twitter, but hashtags and other structural components are used across the breadth of social media platforms, ensuring that findings from this study can be used to inform social media communication practices across a variety of social media platforms.

Methods

Data Collection

Following a data collection approach adapted from Kim et al. (2016) and Gabarron et al. (2014), Tweets containing the term “ovarian cancer” were collected and stored via Tweet Archivist (<http://www.tweetarchivist.com/>), a password-protected online archive. The archive was updated with new Tweets hourly, from May 10, 2017 to May 23, 2017. Data collected included an 18-digit Twitter ID number, user name, time and date the post was made (including universal and local time stamp), language (associated with country code assigned by Twitter based on user settings or country of origin), Uniform Resource Locator (URL) or web address of the profile image, URL of the original Tweet (Figure 1), geographic location, time zone, the Tweet itself (including any links, images, or @mentions), and the number of followers the user had at the time of the Tweet.

Eligibility Criteria

Once data collection was complete, Tweets were imported to a password-protected .CSV (Microsoft Excel) file for cleaning (Figure 2). Data were cleaned via visual inspection. Tweets included in the analysis included Tweets that appeared to be representative of an English-speaking population, based on language code and geographic origin of the Tweets. Location was restricted to the United States to improve the likelihood that the tweets collected would be representative of an English-speaking population, when language code was absent. Location and language were determined by language code attached to the Tweet, geographic location and/or time zone. Tweets with a language code other than EN (English) were excluded from analysis. Tweets published from outside of the U.S. were excluded by identifying any published from a time zone outside of the U.S., then searching the remaining Tweets for any published from a

geographic location outside the U.S. In some cases, an individual Tweet may have originated from a U.S. time zone, but geographic origin was from outside of the U.S. In this situation the Tweet was excluded from final analysis. In other cases, an individual Tweet may only list a geographical location or a time zone, but not both. When one of these pieces of information was omitted, the Tweet was excluded or included based on where the Tweet appeared to be posted from. Tweets that lacked both were included in thematic analysis if they met other inclusion criteria.

Using a method adapted from Wang et al. (2012) and Ferrara et al. (2016), the remaining Tweets were sorted by account name, and each account was entered into Botometer (<https://botometer.iuni.iu.edu/>) (Ferrara et al., 2016). Botometer, formally known as Bot Or Not?, assesses Twitter profiles for account activity, and scores the profile based on how likely it is to be a social bot account (Davis et al., 2016). Botometer assesses highly predictive features, capturing a variety of online behaviors that can be used to distinguish social bots from humans, such as frequency of posting, predictability of content, time of posting, and social networks (Figure 3). Botometer scores Twitter profiles between 0 and 5, with higher scores indicating more bot-like features. For the purposes of this study, accounts that scored 2.5 or lower were labeled human, and accounts that scored of 3.0 or higher were flagged as a bot. Accounts scoring between 2.6 and 2.9 were manually scanned for written content and profile attributes. This manual assessment is described below. For quality control, a random selection of 5% of resulting human accounts and social bot accounts were selected for further assessment. This was the same assessment used for accounts that scored between 2.6 and 2.9. Accounts that were not found or were suspended were also flagged as a bot. Accounts that are suspended or that have been

deleted are likely accounts that have been suspended by Twitter. This is a common quality control practice utilized by Twitter to cut down on spam activity.

As part of the manual check, content and profile images were assessed. Profile images were assessed by conducting Google Image Search or searching for the image over Google using the URL of the image (<https://www.google.com/imghp>). Images that returned duplicate images in search results suggested that the image had been stolen or repurposed from another location online. Profiles lacking a profile picture all together, or those utilizing a repurposed image, suggest that the account is a social bot (Ferrara et al., 2016). Content for accounts that scored between 2.6 to 2.9 and the random selection identified for quality control was examined for repetition (content or post), clickbait content and fear-based language. Clickbait content is defined as text based social media content that provides users with just enough information to peak curiosity but not enough to satisfy it without clicking, attracting attention and accruing link clicks that direct users to specific webpages often heavy with advertisements or malware (Zannettou et al., 2018; Rony et al., 2017). Fear-based language, common across clickbait content, elicits fear by discussing warning signs, using adjectives like “deadly” to describe symptoms in the context of illness (Zannettou et al., 2018; Ferrara, 2015). Tweets posted by accounts that lacked an original image or featured content that primarily included fear-based language or clickbait content was coded as social-bot generated content. Data cleaning resulted in two groups of data: Tweets generated by confirmed social bots and Tweets generated by confirmed human accounts. Both groups of Tweets underwent structural analysis and thematic analysis, described below.

Data Analysis

Using a coding scheme adapted from Chew & Eysenbach (2010), human-generated and social-bot generated Tweets were assessed by first conducting an inventory of each Tweet's structural features and then evaluating what content was expressed and how it was communicated.

Taking an inventory of structural features. The first step, assessing the structural features of social bot-generated Tweets and human-generated Tweets, involved taking an inventory of the physical components that each individual Tweet was comprised of (Lewis et al., 2013; Chew & Eysenbach, 2010). These physical components included 1) images or media (e.g. videos), 2) external links, 3) hashtags, 4) textual content, 5) @mentions and other textual components, e.g. call to action (Figure 1). Images and media were briefly described by the type of content (e.g. photo of girl, animated video about ovarian cancer risk, etc.).

User engagement was also noted. Twitter Archivist records engagement data at the time the Tweet was archived. For the purposes of this study, user engagement included retweets and @mentions. An @mention, or an "at mention," is when the authoring user tags, or names, another user in a post by their Twitter username, in an effort to draw the user's attention to the post or draw the attention of that users' followers. For example, a Tweet that includes @NOCCNational, the National Ovarian Cancer Coalition's (NOCC) Twitter username, is more likely to be seen, and retweeted, by NOCC and their followers, increasing visibility of the Tweet (Zhang et al., 2017). Retweets are reposted or forwarded messages on Twitter. In the data, retweets were indicated by the use of "RT" at the beginning of text. In some cases, retweets had the same text but utilized a different external link. Retweets were coded as unique Tweets, given the variance in external links that accompanied them, and the high percentage of retweets

represented across the data. External links, images, hashtags, user mentions, followers, retweets, and thematic elements were tallied across human generated-Tweets and social bot-generated Tweets and compared using Chi-squared test to test for statistical significance (Campbell, 2007; Richardson, 2011). This assessment described what kind of Tweet components occur together, if any, as well as what structural features are common across Tweets mentioning ovarian cancer and if there is any difference between the structural components used across social bot generated Tweets and human generated Tweets.

Finally, the username and account information for the top ten most retweeted Tweets across social bot and human generated Tweets were examined to get a sense of whether this is a “personal” Twitter account or a “commercial” Twitter account (Lyles et al., 2013). A personal account included any post made by an account that appeared to be managed by an individual. Individual accounts included any Twitter user who had a username not affiliated with any specific organization, place of business, media outlet, or institution, or who did not indicate any such affiliation in their user profile. A “commercial” account included any account that indicated affiliation with an organization, place of business, media outlet, or institution either in their username or profile.

Coding Tweet content. After assessing the structural features associated with each Tweet, the .CSV file was loaded into Atlas.ti and coded for qualitative thematic analysis. Content of the Tweet was assessed to capture the intention of the Tweet—in other words, what was the user’s goal when they shared a particular Tweet? Using an approach to content analysis adapted from Cavazos-Rehg et al. (2016), a subset (10%) of Tweets was randomly selected for review, identifying relevant and recurring themes, generating a codebook to apply to the full sample of Tweets (Neurendorf, 2002). These themes were added to an existing list of codes, developed

based on knowledge of the subject matter. The codebook was then applied to the full data set, making adaptations to the codebook as needed in order to accommodate new themes as they emerged, all the while focusing on what aspect of ovarian cancer is being addressed in each Tweet, the tone used in communicating, and the overall goal of the posting-user. As new codes emerged, Tweets were re-analyzed as needed. When no new categories or themes emerged from the data, the coding scheme was finalized, and the remaining data was coded. To ensure quality control, a second coder coded a different random selection (10%) of Tweets. Percent agreement was used to test for intercoder reliability.

Results

Overview of results

A total of 7111 Tweets were collected between May 10 and May 23, 2017. The median Tweet count was 508 Tweets per day, ranging from 160 to 960 Tweets per day. The Tweets collected during this period included 2830 user accounts, or 2.51 Tweets per user account, of which a total of 32 user accounts were confirmed social bots and 2221 different user accounts posted only one Tweet during the data collection period.

The initial body of Tweets (n=7111) was cleaned, following data collection. Results from this process are summarized in Figure 2. In summary, 91 Tweets were excluded for lacking an English language code, 1178 were removed because they were posted from a time zone outside of the US, and 926 were excluded because they were posted from a geographic location outside of the US. The remaining Tweets were assessed using Botometer, with 2914 receiving a score of 3.0 or greater, 1314 scoring 2.5 or below, and the remaining 688 scoring between 2.6 to 2.9 and saved for further assessment. A manual assessment of profiles rated between 2.6 and 2.9 resulted in a grouping of confirmed human-generated Tweets and confirmed social bot generated Tweets. A total of 36 Tweets from the 2.6 to 2.9 category were determined to be from human twitter accounts, and the remaining 652 from social bot twitter accounts. In total, confirmed social bot accounts were responsible for nearly one-third of the total Tweets collected (n=3566), while confirmed human accounts posted 1350 Tweets. There was a spike in posting activity observed across both groups between May 18 and May 20, 2017 (Figure 4). This observation is discussed in greater detail in the discussion section but may be partly explained by ovarian cancer awareness efforts.

There were 1226 and 1621 user accounts represented across human-generated Tweets and social bot-generated Tweets, respectively (Table 1). Human twitter users averaged 1.1 Tweets per user account, for a total of 96 human generated Tweets per day on average. Just over half (527) of those accounts were responsible for one third of the Tweets collected. In comparison, of the Tweets that originated from confirmed social bot profiles not already excluded due to time zone or geographic location, there were 1621 user IDs present, averaging 2.1 Tweets per user account, and 255 Tweets per day over the data collection period. The top 32 social bot accounts were responsible for nearly one third of the total social bot generated Tweets. The differences observed between number of Tweets generated by each account type, number of user accounts represented between each grouping, and the rate of Tweets per user account were all statistically significant ($p=0.001$, $p=0.022$, and $p<0.001$, respectively) (Table 1).

Structural Components of Human Generated Tweets and Social Bot Generated Tweets

The structural components in both the social bot-generated Tweets ($n=3566$) and human-generated Tweets ($n=1350$) were assessed (Table 2). Over one third of the human generated Tweets (39%) and social bot generated tweets (37%) used a hashtag ($n=522$ and $n=1328$, respectively). This difference was not statistically significant ($p=0.196$). The most common hashtag across the social bot generated and human generated Tweets was the hashtag #cancer. A total of 69% of human generated tweets incorporated a user mention (@mentions) ($n=935$), compared to social bot generated Tweets, 61% of which included a user mention ($n=2182$) ($p<0.001$). There was a statistically significant difference in frequency of the use of external links and images observed across human-generated and social bot-generated Tweets, with 59% of human-generated accounts including a URL ($n=800$), compared to 92% of social bot

generated Tweets (n=3287) ($p < 0.001$). Likewise, the difference between human and social bot-generated Tweets' use of images and media was also statistically significant, 59% (n=800) and 55% (n=1961), respectively ($p = 0.012$). Social bot accounts had an average of 6256 account followers, compared to human accounts, which averaged 9131 account followers ($p = 0.01$) (Table 1).

Frequency and content of Retweets

Social bot accounts were responsible for a higher number of retweets but a similar proportion of Tweets to retweets when compared to human accounts, with retweets accounting for 51% (n=686) of human generated Tweets and 49% (n=1764) of social bot generated Tweets (Table 2). The difference in frequency of retweets observed between human accounts and social bot accounts was not statistically significant ($p = 0.211$). Just over a third (n=227) of the human-generated retweets received one retweet during the data collection period. The Tweet that received the most retweets across human accounts (n=42) encouraged people to learn more about an ovarian cancer talc powder lawsuit. Less than 7% (n=117) of the 1764 social bot generated retweets received one retweet during the data collection period. The Tweet that received the most retweets (n=718) across social bot accounts warned users about signs and symptoms of ovarian cancer, using fear-based language and clickbait to direct users to read the accompanying link. Ten separate twitter accounts were originally responsible for the top ten human retweets, with an average of 476310 account followers. There were five separate twitter accounts originally responsible for the top ten social bot retweets, with an average of 6197 account followers. The difference in average followers of originating accounts for the top ten most retweeted Tweets across human and social bot generated material was not statistically significant ($p = 0.14$) (Table 1). In comparison the average number of followers observed across human accounts was 9131,

and 6256 for social bot accounts (Table 1). The most retweeted Tweets ranged in topic from risk factors, to celebrities impacted by ovarian cancer, to improved diagnostics (Table 3). With the exception of the ovarian cancer lawsuit Tweet, which appeared to come from a social bot account, the top ten most retweeted posts represented in the final human-generated Tweets appeared to be from human-twitter accounts. These human generated retweets contained an @mention, a URL, a hashtag, and an image.

Content of Human Generated Tweets and Social Bot Generated Tweets

Common themes across both human and social bot generated Tweets ranged from event promotion to warnings to research and news dissemination. Just over 10% of human generated Tweets (n=138) used action language, like “learn more” or “be aware,” taking a tone that focused on sharing relevant resources, personal awareness efforts, or increasing awareness about ovarian cancer, compared to just over 13% of social bot generated Tweets (n=473) (p=0.060). Awareness efforts highlighted facts about ovarian cancer, such as “It's #NWHW. Did you know the best way to detect ovarian cancer is by regular women's health exams?” or “#BRCA mutations can increase risk of ovarian, prostate and pancreatic cancers. #PassItOn,” Tweets focused on awareness efforts also promoted specific events, fundraising efforts, or resources intended to draw attention to ovarian cancer as a health issue (Table 4). This included Tweets about studies or trials highlighting study findings or open research recruitment, with nearly 10% human generated Tweets mentioning studies (n=133), compared to almost 2% of social bot generated Tweets (n=68) (p<0.001). For example, one tweet highlighted a clinical trial, stating the following, “RT @HonorHealth: New ovarian cancer clinical trial at VGP Cancer Care Center.” Proportionately more human generated Tweets incorporated a supportive tone (11%, n=149) when compared to social bot generated Tweets (just over 1%, n=50) (p<0.001), and just

over 14% of human generated Tweets (n=190) and almost 6% of social bot generated Tweets (n=209), featured an inspirational tone, defined as Tweets that highlighted stories about beating ovarian cancer and personal accounts from survivors (p<0.001). Patient and survivor stories were a common theme across inspirational Tweets, with nearly 8% of human generated Tweets featuring a patient or survivor perspective (n=106) and just under 5% of social bot generated Tweets (n=176) (p<0.001). For example, one human-account featured a survivor's story, incorporating an @mention and hashtag, Tweeting "Michelle is on @ABC7Chicago this morning as a 3-year survivor raising awareness for #OvarianCancer."

Social bot generated Tweets frequently took on a more somber tone, utilizing fear-based language to convey messaging. Just under 40% of social bot generated Tweets (n=1362), compared to just under 4% of human generated Tweets (n=48), talked about "deadly" symptoms, warning users not to ignore signs of ovarian cancer, such as "Warning: Do Not Ignore These Early And Deadly Signs Of Ovarian Cancer," "7 Warning Signs of Ovarian Cancer #4 is too creepy," and "10 symptoms of ovarian cancer women shouldn't ignore, read and share," (p<0.001). These examples included a call to action, such as "click to learn more," or "learn more today," encouraging users to click on the accompanying link.

Ovarian cancer risk factors, including risk assessment and reduction, were a common theme across social bot and human generated Tweets. Ovarian cancer risk factors that came up in data analysis included talcum powder, genes, and other health conditions. Talcum powder was mentioned in about 4% of human generated Tweets (n=8455) and just over 1% of social bot generated Tweets (n=51) (p<0.001). Tweets mentioning talcum powder, for example "#TalcumPowder Plaintiff Seeks Centralization of Federal Ovarian Cancer Claims Against," referred people to links that appeared to be connected with law offices or legal support resources,

such as @HPCLegalNews. Genetic predisposition and family history appeared in just over 4% of human generated Tweets (n=59) and less than 1% of social bot generated Tweets (n=27) (p<0.001). Finally, other cancers or conditions, including breast cancer, uterine cancer, cervical cancer, prostate cancer, colon cancer, and endometriosis, presented as conditions that were associated with or related to ovarian cancer, were featured in just over 14% of human generated Tweets (n=190) and almost 7% of social bot generated Tweets (n=244) (p<0.001). An example of one of these Tweets is as follows, “Let's remind @acognews that oral contraceptives decrease future risk of ovarian and uterine cancer by 50%!”

Another theme observed across both human generated and social bot generated Tweets was ovarian cancer risk reduction. Less than 1% of human generated and social bot generated Tweets encouraged users to take steps to assess their risk through specific risk assessment tools or screening tools. For example, one Tweet stated the following: “Take five minutes and assess your risk for breast and ovarian cancer. It matters.” Ginger root, however, was a more common risk reduction and treatment topic observed across social bot generated Tweets, with just under 5% of social bot generated Tweets framing ginger as a risk reduction tool (n=170). Comparatively, less than 1% of human generated Tweets referenced this topic (n=11) (p<0.001). An example of a Tweet mentioning ginger is as follows: “Ginger roots kills ovarian cancer & prostate cancer cells better then chemo.”

Other Tweets focused on disease management, from diagnosis to treatment. Ovarian cancer diagnosis was referenced in just over 4% of human generated Tweets (n=57), and less than 2% of the robot generated Tweets (n=43) (p<0.001). Signs and symptoms of ovarian cancer were also discussed across both human and social bot generated Tweets. Just over 9% of Human generated Tweets mentioned symptoms of ovarian cancer (n=126), while nearly 50% of social

bot generated Tweets brought up symptoms and signs of ovarian cancer (n=1696), utilizing fear-based language ($p<0.001$). Ovarian cancer treatment was mentioned in almost 13% of human generated Tweets (n=172), and almost 11% of social bot generated Tweets (n=386) ($p=0.058$). There was also a subset of Tweets that focused on online product sales with an ovarian cancer theme. Many of these Tweets featured a slogan, such as “My Gyn Onc is My Superhero,” along with a product image and link. This theme was more common across social bot generated Tweets, with less than 1% (n=6) human generated Tweets mentioning product sales of this nature, compared to about 8% of social bot generated Tweets (n=283) ($p<0.001$).

Discussion

Principal Findings

This study demonstrates that most content generated by human and social bot twitter accounts consists of retweets covering a range of ovarian cancer related topics. Social media users take to Twitter to seek support, share information, and raise awareness about ovarian cancer. Many users mention ovarian cancer with a tone of encouragement and awareness, relying on personal stories to facilitate mission engagement on behalf of patient advocacy groups and other institutions, or awareness by sharing online resources and the perspectives of survivors. Other Tweets promote resources, including patient specific resources and research advancements, as well as press releases, outlining advancements in ovarian cancer diagnosis and treatment.

The majority of human-generated and social-bot generated Tweets collected during the data collection period included structural components such as URLs, @mentions, images, and hashtags. Previous research identifies these components as variables that contribute to user engagement. Suh et al reported that the presence of hashtags and URLs in a Tweet were positively associated with the probability of a Tweet being retweeted (Suh et al., 2010). Data from this study is consistent with these findings. For Tweets that received the most retweets, it appeared that follower count was also influential across human retweets. The most retweeted Tweets by human accounts originated from accounts that had an average of 476310 followers at the time of data collection. This average is significantly higher than the average follower account observed across the full body of human generated Tweets (9131), as well as social bot generated Tweets (6256). Based on this finding, it appears that follower count of the original posting account may be an indicator of reach of message, a finding that is also supported by Suh et al.

(2010). This observation warrants additional exploration of how to leverage social networks by route of follower count, for the purposes of social network communication. This information may inform social media communication, best practice recommendations about structural features, content, and use of external links in the context of ovarian cancer health communication, while identifying potential gaps or misgivings in the ovarian cancer Twitter conversation.

Awareness campaigns influence twitter activity

Around May 18th, a spike in activity was observed. Based on content of Tweets, this spike was likely due to World Ovarian Cancer Day. World Ovarian Cancer Day appeared to draw people to twitter with their thoughts and resources about ovarian cancer but following this awareness campaign there was a dip in activity observed on May 21. Without viewing website analytics for the landing pages associated with Tweets posted around World Ovarian Cancer Day, it is difficult to say whether or not this increase in posting activity was link to an increase in user action. Additional research is needed to learn more about how a holiday or event may increase not just posting activity, but also information seeking action.

This observation points to an additional shortcoming to collecting data over microblog social media pages, like Twitter. Twitter content is constantly being updated. With this content evolution and themes likely shift. In other words, posts about ovarian cancer collected during one time period will vary if compared with another sample of tweets, collected over a different period of time (Lafferty & Manca, 2018). Shifts in themes may be driven by awareness events, holidays, or current news. Future research should focus on implementing additional Tweet collection periods to combat this limitation, applying the same methods to analyzing other data sets collected at different times.

Differences in Tweeting

It was clear that social bots are responsible for generating more content per day and per account, when compared to human-operated Twitter accounts represented in the data set. Social bots posted twice as many Tweets during the data collection period. This observation aligns with prior research (Benevenuto et al., 2010; Lee et al., 2010), indicating that one of the markers of social bots is a higher posting frequency, when compared to human user accounts. Human generated content varied from social bot generated content in other ways, too. Human generated Tweets featured research and support related content, as well as inspirational messaging at higher rates than social bot generated content did. Human generated Tweets also mentioned disease management and risk factors more often than social bot generated Tweets. Social bot generated Tweets, on the other hand, mentioned signs and symptoms of ovarian cancer at much higher rates, tending towards fear-based language more than human Tweets did. In this context, social bots leveraged the public's fear of disease combined with a call to action encouraging users to visit a website, driving traffic to a specific external link. These external links take users to websites that are easily monetized through the use of advertisements. This is a well-documented tactic employed by social bots (Shao et al., 2017). Likewise, social bot accounts were also more likely to link users to specific products than human generated Tweets. Future research should explore the impact of social bot generated content, to discern whether or not such material is treated as just noise, or if it has an actual impact on the type of content users interact with.

The challenge of detecting social bots

A major hurdle in completing data analysis in this study was discerning human generated Tweets from social bot generated Tweets. Earlier iterations of the proposed cleaning mechanism proposed identifying social bots by seeking accounts with a higher than expected Tweeting

frequency. Following this step, 10% of the remaining accounts were randomly selected and assessed manually. Most of this random selection was discovered to be social bots, via Botometer. When Tweets were examined more closely, it appeared to be quite common for accounts with a similar user name, for example EtsyLovely and EtsyFabulous, to post the same Tweet at the same time. Since these accounts were responsible for only one or two Tweets in the data set, the low frequency made it appear to be a human-operated twitter account based on frequency alone. However, when these types of accounts were assessed by Botometer, they either had been suspended since data collection, indicating that they were a social bot (Freitas et al., 2015), or they scored high in Botometer based on other social bot criteria. In other words, as social bots become more sophisticated, they get more difficult to detect through rudimentary means (Ferrara et al., 2016).

Previous research supports the observation that more sophisticated bots can slip through less advanced screening mechanisms (Ferrara et al., 2016). While human detection continues to be a reasonable approach to identifying social bots, this approach is self-limiting since it requires significant manpower to manage large sets of twitter data (Ferrara et al., 2016; Wang et al., 2012). Social bot detection software offers another solution, using a detection algorithm to discern bots over human accounts (Elyashar et al., 2013). However, this type of detection software leaves a middle ground of accounts that share some traits with social bot accounts, and some with human accounts, making it an imperfect solution as well. For example, Botometer scores a number of accounts in the middle of the 0-5 scale, 2.6-2.9, accounts that require manual assessment. In the end, a combined approach for social bot detections was the most appropriate for this data set—first screening accounts for frequency, then screening accounts using a social bot detection tool, and finally manually assessing any accounts that scored between social bots

and human accounts. By using the combination of social bot detection software and manual assessment of accounts scoring between 2.6 and 2.9 via Botometer, social bots were easily identified, making it possible to focus on human-generated content.

The initial goal of this study was to focus on human generated content, but both human and social bot generated Tweets were analyzed. This decision was made in response to the observation that the two types of accounts are getting more and more difficult to distinguish, by Twitter users and public health researchers, alike, without significant foot work. As social bots continue to increase in complexity and sophistication, however, it is likely that if tools like Botometer don't grow along with them, they may fail to screen effectively for social bots. This is a concern for any researcher interested in mining twitter data to learn more about how the public interacts with, and produces, health information over social media. It appears that qualitative methods drawing on crowd sourcing and manual assessment will continue to be an important tool for screening and quality control. Alternatively, if social bot content is indistinguishable to both social media users and researchers, perhaps the more important question is how, and if, social bot generated content influences online health information seeking.

Limitations

One limitation to this study was the inability to compare engagement data. Prior studies have focused on assessing how structural components impact engagement and sharing, but the data collection tool used for this study made this difficult since engagement was only updated at the time of collection. Without assessing specific Tweets from the moment they were posted (for example, perhaps a retweet was originally posted before the data collection period commenced), and onward over a designated period of time, adding engagement data over time, it is difficult to get a true estimate of engagement over time to compared different Tweets to one another. Some

of the most retweeted Tweets collected during data sampling, excluded the original Tweet, since it was posted outside of the data collection period. This along with subtle changes to text in Tweets where the same link was shared, made measuring an exact rate of engagement of time difficult. Another limitation was the data collection period. Data collection revealed that social media mentions of ovarian cancer increased during and around World Ovarian Cancer Day. As a result, the ovarian cancer mentions during the data collection period may have been generally elevated compared to the rest of the year, and data collected may not be representative of other time periods. Future studies should include additional data collection periods outside of awareness events.

Finally, the use of third-party data collection and analysis tools, like Tweet Archivist and Botometer, present possible limitations. There are different versions of tools that archive social media posts and provide some rudimentary assessment available. Each third-party tool completes their proclaimed tasks differently, based on a different set of algorithms. These tools, which are often homegrown, are not subject to outside quality control and may not be updated frequently. This is particularly problematic for tools like Botometer, whose efficacy depends on frequent updates, in order to keep ahead of evolving social bots. If programs like these are not updated on an ongoing basis, they may fail to perform their intended function accurately, impacting the quality and accuracy of data collection and analysis. This was not such a concern for Botometer, which is the product of a collaboration between a number of academic institutes and is updated frequently.

Tweet Archivist is one of many fee-based third party Tweet collection tools. Other Tweet archive tools may include different data points or may manage components like retweets, differently. For example, Tweet Archivist linked retweets directly to the original Tweet, rather

than the retweeted one, making direct engagement observation difficult. Another potential issue with the use of third-party data collection tools is that some are limited to a certain number of tweets or a certain time period. For example, other services may only record the last 2000 Tweets or only those posted in the last week. Tweet Archivist does not pose this particular limitation but depending on the researcher's goals and what data points they require, another third-party data collection tool may be more appropriate, and without frequent updates or maintenance, Tweet Archivist may encounter quality control issues.

Aim 2

Introduction

Online information seeking: risks and benefits

People frequently turn to the internet for health information before consulting their healthcare provider (Volkman et al., 2014; Fox & Duggan, 2015;). This trend is forecasted to increase with the rapid growth of mobile health applications, mobile device ownership, improved access to personal health information via electronic medical records, and the general availability of online health services (Mills et al., 2015; Carlsson, 2009; Abrahamson & Fisher, 2007; Rutten et al., 2005). As the availability of online health information continues to grow, the demand for online patient resources related to cancer genetics will increase along with it.

There are problems with how the internet provides information to people. It is the public's ability to freely engage with and exchange resources over the internet that makes health information more available, regardless of information accuracy or quality (Wong et al., 2013; Miller & Bell, 2012; Zulman, et al., 2011; Arora et al., 2008; Eysenbach et al., 2002). Studies exploring the quality of online cancer information draw attention to the lack of accuracy and quality assurance across online health information sources (Banki et al., 2016; Sobota & Ozakinci, 2015; Lawrentschuk et al., 2012; Kukreja et al., 2011; Friedman et al., 2011). Information available over social media sites may be plagued with misinformation and devoid of citations (So et al., 2016; Abramson et al., 2015; Cole et al., 2016), and health information posted to commercial websites is often more focused on selling a product or service over providing non-biased information (Grewal & Alagaratnam, 2013). These observations spark concern about how inaccurate health information may mislead decision making, encourage

harmful health practices, or damage the physician-patient relationship (Ventola, 2014; Lau, et al., 2012; Murray, et al., 2003).

These concerns prompted the development of a list of recommendations for health information seekers from the National Institutes of Health (NIH) (National Institutes of Health, 2011). These recommendations encourage users to note the following details, among others, when visiting a website in search of health information:

1. Who runs the website?
2. Who funds the website?
3. What is the purpose of the website?
4. Does the website include citations or indicate original sources?
5. How does the website list or note references?
6. Is the content up to date?

The NIH's recommendations acknowledge that even though the internet has made information more readily available, it also enables the rapid dissemination of false and misleading information. The goal of these recommendations is to help the consumer better discern quality information from false information online, vetting online sources.

Online health information has both positive and negative potential effects. Recent meta-analyses suggest that online health information does not negatively impact patient outcomes (Nolke et al., 2015; Crocco, 2002), and instead may benefit cancer information seekers. At its root, the internet improves an individual's access to health information. Improved access to health information benefits the individual in a variety of ways, facilitating patient engagement and awareness (Lemire et al., 2008), supporting coping skills (Van Der Molen, 1999; Derdarian, 1986), reducing uncertainty and distress (Ress & Bath, 2001; Fallowfield et al., 1990;

Derdiarian, 1986), empowering the individual to make informed decisions about their medical care (Lemire et al., 2008; Dutta-Bergman, 2004), and improving patient satisfaction with healthcare decisions (Luker et al., 1995). Cancer patients increasingly turn to the internet with their questions about disease diagnosis, prognosis, and treatment, using this information to shape and inform conversations with their provider (Huerta et al., 2016; Fox & Duggan, 2015; Volkman et al., 2014; Hong et al., 2012; Fox, 2011). They benefit from being able to communicate with others going through similar health situations and being able to access cancer advocacy organizations online (Maddock et al., 2011).

The online information seeking behaviors of individuals at risk for ovarian cancer

Current research regarding the information seeking habits of cancer patients and survivors, the primary targets of cancer information providers, is robust, but little is known about the information seeking behaviors of individuals with a genetic predisposition for cancer, whose informational needs differ (Rutten et al., 2016; Roach et al., 2009). Differences in informational needs and goals modify the information seeking process (Johnson et al., 1995). For example, an individual battling cancer may experience a greater sense of urgency for information, compared to an asymptomatic person with a genetic predisposition (Kernisan et al., 2010; Noh et al., 2009). The majority of cancer patients (74-98%) report that they want as much information as possible about their condition, but the percentage who actually seek cancer information appears to be dependent on factors unrelated to their diagnosis (Rutten et al., 2005; Chen & Sui, 2001; Foltz & Sullivan, 1996). These variables may include demographics, exposure to a friend or family member with cancer (Carlsson, 2009; Czaja et al., 2003; Rees & Bath, 2001), or their perceived risk of developing ovarian cancer in their lifetime (Wong, 2012; Weaver et al., 2010; Kaphingst et al., 2009; Roach et al., 2009).

Perceived risk is defined as one's belief regarding how likely personal harm is given a particular risk factor (Weinstein & Klein, 1995). Perceived risk may be affected by any number of variables, including genetic risk information (Aspinwall et al., 2014). For example, receiving a positive genetic test may increase an otherwise healthy individual's perception of risk, and impact their motivation to seek information (Wong, 2012). That individual may be more or less inclined to access, consume, or trust certain online sources of health information (Aspinwall et al., 2015; Wong, 2012). More research is needed to better understand how individuals with a family history of, or genetic predisposition for, ovarian cancer address health-related questions and engage with health information online. This study seeks to address this knowledge deficit by directly observing the online information seeking behaviors of individuals who have been told by a health care provider that they have an increased risk of ovarian cancer due to a defined genetic predisposition or family history.

Study Aims

Using the Comprehensive Model of Information Seeking (CMIS), developed by Johnson (1997), this study directly observed the online information seeking behaviors of individuals who have been told by a health care provider that they have an increased risk of developing ovarian cancer. The CMIS model has been widely used to explain the information seeking behaviors of cancer information seekers, drawing on concepts and findings from a synthesis of the Health Belief Model (Rosenstock, 1974), Uses and Gratification Theory (Rubin, 1986), and the Media Exposure and Appraisal Model (Johnson et al., 1993A & 1993B) (for an extensive discussion, see Johnson, 1997). Findings from this study may assist in improving public health professionals' understanding of how patients use the internet to address health concerns, using

this insight to improve access to quality health information, with the goal of ensuring that the right people, receive the right information, at the right time.

Methods

Overview

This study used a qualitative approach, driven by Johnson's Comprehensive Model of Information Seeking (CMIS), to better understand the information seeking behaviors of individuals who have been told by a healthcare professional that they are at an increased risk of developing ovarian cancer, due to a confirmed genetic predisposition or personal or family history of hereditary cancer. This study drew on a series of questionnaires and contextual, task-driven interviews with participants, designed to understand online health information-seeking behaviors. A total of 12 interviews were conducted with English-speaking women over the age of 25. Eligibility criteria included individuals who lived in the US, have access to a computer with an internet connection, have at least one ovary, have not been diagnosed with or treated for ovarian cancer, and have been told by a healthcare professional that they are at an increased risk of ovarian cancer.

Participants and sampling

Participants were recruited through clinician referral and over Facebook posts. Healthcare professionals located in King County, Washington, with a focus on clinicians working in gynecological oncology and genetic counseling, were approached with information about the study over email, in person or via phone. This communication included a brief explanation of the study and the eligibility criteria for the clinicians themselves, as well as a drafted email and flyer they could share with patients who might be a good fit for the study. A study flyer was also posted online over Facebook.

Interested parties were prompted to contact the primary investigator directly by email. Eligible participants include women over the age of 25, who live in the United States, speak

English, have access to a computer with an internet connection and have at least one ovary. Participants must have been told by a health care provider that they are at an increased risk for ovarian cancer due to a pathogenic gene variant, or personal or family history of breast cancer, or family history of ovarian cancer. A confirmed genetic predisposition would include a documented mutation in one of the following eleven ovarian cancer genes: *BRCA1*, *BRCA2*, *BRIP1*, *PALB2*, *RAD51C*, *RAD51D*, *BARD1*, *MSH2*, *MSH6*, *MLH1*, or *PMS2*. If the participating individual cannot produce the name of a specific cancer gene but indicates that they were told by a clinician they are at high risk for ovarian cancer due to a genetic predisposition, an ovarian cancer susceptibility gene, or due to a family history of breast or ovarian cancer, they were invited to participate if they were otherwise eligible. Anyone currently being treated for, or previously diagnosed with, ovarian cancer was excluded from participation.

Prior to enrollment, participants completed a brief eligibility questionnaire over the phone. Eligible individuals scheduled a time to complete an interview and received an electronic consent form via REDCap. Although small, the sample size was deemed adequate due to data saturation (Glaser & Strauss, 1967). Data saturation was reached when further coding was not feasible and when interviews ceased to yield new information and any subsequent interviews would likely fail to yield any additional information or coding categories (Guest, 2006).

Data Collection

This study used a context driven approach to directly observe the online health information seeking behaviors of those who have been told by a healthcare professional that they have an increased risk of ovarian cancer due to family health history or a confirmed pathogenic genetic variant. Interviews ranged in duration from 30-75 minutes. Interviews took place over a video conferencing service that allowed screen share, e.g. Skype, Google Hangouts or Zoom,

between April 2017 and February 2018. During the interview, participants were seated at their personal computer in a location of their choosing. The interviewer initiated a video conference call at an agreed upon time from a remote location. Each interview consisted of two sections. Section one included a questionnaire adapted from the eHealth Literacy Scale (eHEALS) and a brief series of questions intended to gauge the participant's relationship or personal experiences with ovarian cancer. Section two of the interview used a qualitative approach to observe information seeking behaviors in real time. These are described in greater details below (Figure 5).

Participant responses to section one of the interview were manually recorded on paper. Data for section two of the interview was gathered using a method described by Kushniruk & Patel (2003), using a screen-capture tool, e.g. Panopto. Screen capture software captures a video recording of the user-computer interactions and an audio recording of the conversation. The end result was a single video recording with audio for each interview. By conducting interviews over a video conference service, data were gathered remotely, while the participant responded to prompts where they would most likely be to naturally carry out the assigned task, for example on their computer at their home or work place (Holtzblatt, et al., 2005; Raven & Flanders, 1996). Unlike other approaches that rely on participant report and recall, this method enabled researchers to directly observe information seeking behaviors in real time (Blomber et al., 1993).

Quantitative measures

Section one of the interview focused on measuring components of the CMIS model using the adapted eHEALS questionnaire. The adapted eHEALS questionnaire included 12-items. The eHEALS questionnaire was originally developed to quantitatively measure an individual's combined knowledge, comfort, and perceived skills at finding, evaluating, and using electronic

health information (Norman & Skinner, 2006). The adapted questionnaire asked participants to reflect on the internet and social media as a source of health information. Participants responded to questions using a Likert scale from one to five, where higher scores indicated greater agreement or a greater sense of importance or usefulness in regard to the referenced resource or online information seeking action. Section one also included seven additional questions, shaped using the CMIS model, intended to gauge the participant's relationship or personal experiences with ovarian cancer (Table 5). Participants were asked to rate their agreement with statements about ovarian cancer using a scale from one to five, where higher scores indicated greater agreement with the statement.

Qualitative Data Collection

In Section 2, participants used their personal computers to respond online to a series of scenarios, while the interviewer observed the participant's information seeking in action, learning about their goals and perceptions of online health information sources (Feather, et al., 2016; Kantner et al., 2003; Raven & Flanders, 1996). Participants were presented with one scenario at a time and asked to respond to the information seeking task online using a search engine of their choice, in a manner that would be the most natural to them. Each task required users to use the internet to locate a specific piece of health information or address a health concern, related to ovarian cancer. Task scenarios, reflective of the information needs of individuals who are concerned about cancer, included the following hypothetical situations: experiencing symptoms of ovarian cancer, considering family history of ovarian cancer, implications for family members, and genetic risk factors for ovarian cancer (Metcalf et al., 2009) (Table 6). During the interview, the researcher asked follow up questions to help assess antecedent factors, or underlying beliefs about specific information carriers, while examining

information carrier characteristics important to individuals with a confirmed genetic risk of ovarian cancer. Remote interviews allowed participants to complete each information seeking task using their own computer and in the location they would be most likely to naturally complete the tasks. It is possible that their responses would have differed without the direct observation of the interviewer, reducing the Hawthorne effect, however the demand for follow up questions and the use of the think-aloud approach described below, made this level of independent interview completion difficult (McCarney et al., 2007).

The interviewer used a “think-aloud” approach as the participant responded to each task (Beyer & Holtzblatt, 1997). The think-aloud approach required participants to talk through their thought process, discussing what they are seeing and considering as they work (Makri et al., 2007; Charters, 2003; Kushniruk & Patel, 2004). A series of follow up questions were used throughout the interview, prompting the participants to expand on their thought process and clarify information-seeking actions observed during the course of the interview (Table 7). This approach reveals motivations behind computer-human interactions, allowing the researcher to use follow-up questions to gain a better understanding of the information seeking behaviors exhibited (Case, 2012; Kushniruk & Patel, 2004). The digital recording of all interviews was transcribed verbatim by the researcher.

Analysis

Descriptive statistics were used to summarize responses to the adapted eHEALS and accompanying questionnaires. Next, video recordings from each session were analyzed in Atlas.ti, a program well-suited to accommodate the coding of video data. The transcription was analyzed using content analysis, where video data was assessed as supplementary data, or visual data that complimented the transcribed audio, demonstrating accompanying actions (Konecki,

2011). Using an approach adapted from Konecki (2011), each video recording was assessed in five-minute intervals for engagement, defined as clicks or another interaction with page content, and any accompanying audio. The result was a brief description of what happened during each interval. Characteristics of the respective online source's participants referred to over the course of the interviews were also noted. Familiarization with the data led to the identification of initial codes, guided by the CMIS model. Codes were refined and clustered into categories and the resulting codebook was used to code each transcript and accompanying video data. Audio not associated with any engagement was noted separately and coded.

Coding was completed by two separate coders, using thematic analysis and a constant comparison approach to coding (Glaser & Strauss, 1967). Throughout the data collection and analysis process, a detailed audit trail was maintained to ensure accurate research procedures, record methodological or interpretive insights, and facilitate the writing process (Charmaz, 2006; Leininger, 1994). Once coded, all the resulting codes and notes were reviewed together to identify common categories or unique concepts that appeared across the videos. In reviewing data from all the recordings alongside one another, the goal was to get a sense of what factors, if any, reemerged across the interviews.

Results

Principal Findings

In section one of the interview, summarized in Table 8, all participants reported that they had either a family member or a friend who had been diagnosed with breast or ovarian cancer. The majority of the participants agreed they were at high risk of developing ovarian cancer in their lifetime, with an average agreement score of 3.5, where higher scores indicate greater agreement. Likewise, the majority of the participants also agreed that their genes put them at higher risk of developing ovarian cancer, with an average agreement score of 3.9. The majority of participants also agreed that ovarian cancer was a devastating disease, reflecting their personal beliefs concerning the severity of the disease, with an average agreement score of 4.6. Participants also reported that they believed they could reduce their risk of ovarian cancer by engaging in preventative behaviors, such as healthy diet and exercise, with an average agreement score of 4.4, and said they knew how to find information about ovarian cancer, with an average agreement score of 4.1.

All interview participants reported that the Internet was a useful source for ovarian cancer information, that was important for them to be able to access for health information, with an agreement score of 3.8 and 4.7, respectively. Participants did not feel the same way about social media as a source for health information, with most reporting they did not think it was useful to get health resources from social media, nor was it important for them to be able to get resources over these platforms, with an average agreement score of 1.8 and 2.1 respectively. On average, participants reported that they knew what kinds of health resources were available online (agreement score of 3.8), where and how to find them (agreement score of 4.1 and 4.2, respectively), how to use them to answer questions about their health (agreement score of 4.4),

and how to use these resources to help themselves (agreement score of 4.2). Each participant questioned the reliability and accuracy of some of the online information reviewed throughout the course of the interview, but they also reported feeling confident in their ability to discern high quality health resources from low quality health resources online, with an agreement score of 4.2, and said they had the skills they needed to evaluate online sources (4.1). Finally, participants also reported that they were confident in their ability to use online information to help make health decisions, with an average agreement score of 3.6) (Table 8).

In section two of the interview, participants responded to the task scenarios. During this section, all of the participants opted to start their response with Google. Participants also tended towards more familiar sources, but when faced with an unfamiliar source displayed a relatively consistent method for assessing the quality of this source. Almost every participant explained that they would use information they found online to inform conversations with their health care provider. Only one participant mentioned using a specific social media platform when responding to the scenarios. These results are discussed in greater detail below.

Relying on Google

All participants opted to begin their online health inquiries with Google as their search engine of choice during the interview. They used a consistent combination of search terms for each information seeking task, including “genetic risk,” “ovarian cancer,” “breast cancer,” “BRCA,” or “ovarian cancer symptoms.” One participant said:

“...the first thing I'm thinking about is what my search terms should be to get useful information about the topic. Finding my trusty Google... I'm just going to start with a general Google search and see what I end up with.”

Participants scrolled down through the results, reading the search results and looking for key words related to their search. Participants looked at the details under each search result, to assess

the initial search results and visited individual websites to further assess relevance. Search results always started with sponsored results and ads, which included a symbol with the word “Ad” in front of the link to the website, followed by a listing of organic results. “Organic” results in this context are defined as search results that were not paid for. Each individual search result included several pieces of information about the website. Participants referred to each of these components when reviewing search results to decide which information source would best meet their needs. Components included the title of the page in blue font, the URL of the page in green font, and a brief description of what the page is about. Participants looked for key words related to their information search when first reading these components. They also focused on the first page of search results, and rarely went beyond these first hits when performing the information seeking task.

When a participant found a search result that they felt met their needs, they typically clicked on the result and went to the website to further assess. Analysis of this initial review of search results revealed consistent themes in how users identify online information resources that they trust and how they assess the quality of online information and what users do once they find that information. This assessment process is described in greater detail below.

Familiar sources and top-level domains get the click

Before clicking any webpages for more information, in addition to looking for key words participants scrolled through the top search results, looking at two specific characteristics. These characteristics included name recognition or source familiarity and the top-level domain (TLD) utilized in the website address. Source familiarity was one of the most common characteristics mentioned across all interviews. Participants noted the authoring organization behind each online resource, looking for resources originating from organizations or institutions they were familiar

with and trusted, as the result of previous experience or familiar branding. Sources that were unfamiliar drew questions, for example, one participant said the following when reading search engine results:

“Um, Ovarian Cancer Research Fund alliance. So now I'm thinking about, well how reputable is this source? It's a .org, so it's not like a government source or anything like that. So like, who are these people? Um.”

The identification for the authoring organization was sometimes provided in the description of the webpage, but other times participants referred to the website URL to determine the name of the authoring organization. Patient advocacy groups, cancer treatment centers, or other cancer support resources, that they recognized either as the result of geographic proximity, referral from a friend or family member, or an organization they had personal experience with, all warranted noting or prompted additional investigation. Common familiar sources included patient advocacy groups, such as Susan G. Komen, and cancer treatment centers, such as Cancer Treatment Centers of America. Familiarity was often anchored to simple name recognition, or lack thereof.

For example, one participant said:

“Susan G. Komen, I mean that, that has always been a very familiar thing throughout. Then I find some that I might definitely pass by. I mean, just things that are unfamiliar, I guess. Like this one, kind of lower down here. NCBI, it's just kind of ambiguous. Just don't really know what that is from or for.”

Another explained that they preferred names they had heard before from exposure or referral, stating:

“I like to go with the big ones [health resources] that you know about. The places that you heard of. If it's been in the news. Or, someone who's been sick in life and they talk about the different resources they have, because I feel like they'll have a better idea.”

Other times participants noted that they knew about a specific organization because the brick and mortar version of the online organization, or physical presence of the organization, was located in close proximity to their home.

When participants identified a resource that was unfamiliar to them, they relied on the web address itself, specifically the top-level domain name, or TLD, to assess the trustworthiness of the online resource. The TLD is the last three letters at the end of a website address and is one of the domains at the highest level of the hierarchical Domain Name System (DNS) of the internet. Some common examples of a TLD include .com, .org, .edu, or .net. Preferred TLDs, or those that participants indicated were more reliable, or more likely to be sources they trusted, included .org, .edu, and .gov. Alternatively, several participants indicated that webpages that ended in .net were less likely to be webpages they would seek health information from. One participant stated:

“I don't know about cancer.net. I'm not sure about that. Cancer.org. Hmm. [Pause] I think I would start with the CDC.gov.”

Geographic origins of online resources are important to information seekers

Several participants also noted the country code of the web address. Country codes are another type of TLD, represented by a two-letter combination in the web address. Three participants noted international sources in their search results and indicated that they were interested in seeking health information from outside the United States. Participants expanded on this interest, explaining that they believed other countries may be making new strides in research, identifying new approaches to ovarian cancer diagnostics and treatment, and that this type of information may be otherwise unavailable in the US, even though it may be something that could benefit them. One such participant reflected on why research from the UK was of interest to them:

“Because I do think that there are things going on in different parts of the world where they are making gains perhaps that I just, I would love to be acknowledged and most of them I do. If I did talk to a doctor, I could talk to them about different, um, different advances that have been made. For example, I was just listening on public radio about preeclampsia with women who are pregnant. and that the United States has the highest

amount of women who tend to umm to tend to have complications from preeclampsia or death and there are other countries who have figured it out. How to make sure that there are the complications from something like preeclampsia that which certainly is treatable. So I do like to find out what other people are doing to make sure we opened our eyes to other, other possible strengths that other countries are doing.”

Other participants used geographic origin to weed out sources that they said they did not think was applicable to them or their situation. This was specifically an issue for participants who ended up selecting research publications in response to interview prompts. If participants said that a study population reported on was not representative of them, for example if the study population was a different race or ethnicity or from a different part of the world, the participants indicated the results reported in research publications may not be relevant to them. One participant, when reviewing findings from a study about women from an international location, said the following:

“I don't think that I could generalize that study to my life...there are so many environmental factors there.”

Users conduct quality assessment of online resources

After looking for familiar sources and scanning the URL, participants selected the source they thought would address the information seeking task at hand. On average, participants clicked on one to three sources per search. Once they landed on the webpage, participants examined the utility of the page after they had judged that it met their initial criteria and clicked on the page. If the source was familiar, they briefly scanned the page to see if it addressed their information seeking task. If the information source was unfamiliar, however, participants engaged in a more rigorous process to verify the source, gathering more information about the webpage and authoring organization, noting the usability of the resource, and seeking consensus across multiple source as a quality control measure.

Participants sought out the “About Us” page, when investigating a resource that they were not familiar with. For example, one participant said the following when seeking more information about one webpage:

“Clearly, they are passionate about ovarian cancer, they are talking about saving women's lives, but I would want to know who these people are. Um, [reading] there's “about us,” alright, who are you people? About you. Who are you? [reads] You have a mission. That's nice.”

The “About Us” page was found under varying titles, but was typically present and easy to find, and provided participants with information about the authoring organization behind the website.

Participants also noted whether or not the webpage seemed to be “trying to sell something,” or noted potential conflicts of interest, favoring organizations that were conducting research over commercially based organizations, that participants felt may impact the quality of the information provided. For example, one participant noted:

“I mean generally, I think that I trust nonprofit organizations, or government organizations, or whatever, kind of more than I would trust somebody that's out to sell ads or be sponsored by people.”

Another participant mentioned that they would avoid health related websites that appeared to be sponsored by a pharmaceutical company, expressing concern about the websites ability to provide objective health information. The participant explained:

“I would have to look at it further to see if... it is sponsored by a pharmaceutical company. So, if it is sponsored by a pharmaceutical company, I would probably disregard it [the source], or put it at a lower priority, because I think that they're selling their therapy rather than being objective about the whole situation.”

Several participants discussed the importance of website usability, or how user friendly the website interface was perceived to be by the participant. Participants noted things like how easy it was to read, focusing on font size, type, color, or contrast, as well as content organization. Users expressed a preference in font that was easy to read, that utilized high contrast. One

participant talked about whether the page “cut straight to the point,” noting that they wanted to be able to figure out at a glance what this page was talking about, favorably commenting on infographics. One participant mentioned the use of infographics in particular, stating:

“Ooh, a schnazzy infographic. That could give me more, some useful information.”

Participants who discussed usability indicated that high contrast text and certain fonts made text easier to read, noting that easy to read font increased their comfort with the source, while “Comic Sans” was a “deal breaker.” Another participant noted that the appearance of a webpage was important when looking for a source they could trust, explaining:

“I guess if it looks like the website is a quality website, and not just you know, something that was kind of hacked together. Yeah, if it's formatted well, and has nice fonts, and things.... You know, just like aesthetically nice-looking website, and images that are helpful to go along with the text, as well. I mean, and not like really graphic images, I would say, but you know, more kind of along the lines of almost stock imagery that makes sense along with things.”

One participant noted that when they open a new webpage, they looked to see what types of imagery the page used, specifically if the included images added something to the page, if they paired well with the text, and if they appeared to be stock images, favoring candid images relevant to the page material. Participants also looked to see when the page was last updated, scanning the website for a date. Several participants noted “professionalism” as an important factor, favoring online resources that looked like they had been professionally designed as opposed to webpages that appeared more home grown or outdated. One participant also noted however, that given advancements in technology, a professional “appearance” could not be the only assessment criteria for an online resource, adding that:

“I’m always looking for presentation, professionally done, when was it last updated, quick links to everything, or references that I can explore further.”

Another participant also commented about professional appearance being an important factor in building trust, stating:

“..nowadays anybody can make a website and a lot of them look very professional and fancy and it can be a challenge to figure out, you know, um, whether they are reputable sites.”

Participants also expressed an interest in interactive webpage components, in particular quizzes that estimated cancer risk where users were invited to enter information about their specific risk factors and behaviors and find out what their estimated risk for developing cancer was based on the information provided. An example of one of the quizzes several participants referred to over the course of the interviews can be reviewed in Figure 6 (<https://www.assessyourrisk.org/>).

Participants suggested that quizzes were an acceptable way to assess risk, and reported that they trusted this type of information as a starting point in their information seeking.

Participants also noted whether or not the website they traveled to looked like a “forum.” An Internet forum is a type of social media where users can hold conversations and respond to posted questions and messages. Participants discussed how a forum differed from a curated informational website or online resource, reflecting on the appearance of the website adding that forums often looked less professional. One participant said:

I don't trust forums at all really, any kind of mom forums, or medical forums, or anything that ... I try to avoid them as much as I can, and I don't know that this is ... it looks like it's just a lot of people commenting about it. Yeah, I would probably not get lost in all of that stuff, so I'd probably get right out of there.

In the context of health information, online forums may be condition specific, where users ask questions about symptoms, treatments, prognosis, or seek general support. Almost every participant mentioned a web forum or online support group, but participants appeared to have mixed ideas about the role such a resource would play into their information seeking endeavors. Most participants stated that these types of websites would either not play a major role in their

information seeking or they would avoid these types of resources. One participant did say they would consider going to a support group or a health message board for information about a symptom, to reflect on their personal experiences with others who perhaps have experienced a shared type of symptom or pain.

The majority of participants noted one final step when assessing quality and accuracy of online information—consensus. Consensus seeking was particularly common when participants were prompted to search for information about ovarian cancer symptoms. They stated that they would seek information across multiple webpages, to get a sense of how the majority of these online resources answer their specific question. If they saw several sources associating a specific symptom with ovarian cancer, they may be more inclined to schedule a visit with their doctor.

One participant said:

“If you read any of the [online] articles, you would look for recognizable sources, look for consensus across different sources.”

Participants who talked about the importance of consensus indicated that once they start seeing a theme across their online searches, such as a symptom of ovarian cancer, certainty over that information would increase. All interview participants responded to this prompt, but most indicated that they would most likely not look up information about disease symptoms online.

Participants did not express any reluctance to looking up risk information or details about hereditary cancer. One participant said:

“[If you look up symptoms online, the results will tell you either] ate a marble or you have ovarian cancer.”

Another participant stated:

“Google can lead you down a scary road.”

Online resources inform conversations with health care providers

After assessing the online information source, participants were invited to reflect on how they might use the information. Every participant noted that they used online information to identify questions to take to their healthcare provider and inform that discussion. One participant said:

“The internet can be a good place to start learning, then you would take all that information to the doctors to talk and make a plan. I’m the most important person to me. They (the doctor) have lots of people. I need to be the most informed person.”

Another participant added that:

“The internet is a great place to take your initial questions and a path to find questions to bring to your doctor. It can help you brainstorm what kinds of things you should be asking.”

Participants with personal experience receiving genetic test results that informed their ovarian cancer risk talked about the time lapse between receiving a positive genetic test result (i.e. finding of a pathogenic variant), and their follow up doctor’s appointment with a specialist to discuss their results and what they meant, noting that there was often several months in between. These individuals said they used online resources to help them identify and articulate their questions, enabling them to use their appointment with their specialist more efficiently, bridging the information gap between risk identification and follow up. One participant said:

“It’s difficult to get ahold of your doctor, may not be able to easily get an appointment. You’re likely to go three months from diagnosis to follow up. Pelvic pain or ovarian cyst? [The] Internet may be a great way to get some information, to help you make the most of the time you have with your doctor.”

Another participant added that it would be nice if their doctor recommended specific online resources to help them address some basic questions to help prepare for their follow up visit. One participant was less interested in using the internet during this time. They talked about an “in between stage,” or the time between getting a positive genetic test result and when they had their

follow-up visit with their healthcare provider. They said during this time it was best to “stay off the internet,” citing unnecessary worry.

Participants noted that they would also use online information to assess eminent risk and identify when a trip to the doctor may be necessary. Participants noted that they might use online information about worrisome symptoms to decide how quickly they needed to schedule a visit with their doctor, with several noting that in addition to this information they also might just “wait and see,” to see if the symptom resolved itself without requiring further intervention.

Social media as a family communication tool

Most participants did not reference social media over the course of the interview at all. The one participant who did, talked specifically about how they had used social media to inform estranged family about their genetic test results. They specifically described how they used social media to create a closed Facebook group. The participant said they invited family members over Facebook to join and shared material they received from their doctor about their confirmed pathogenic genetic variant.

Discussion

Most online information seeking behavior research utilizes research methods that rely on participant recall and perceptions of behavior, such as surveys (Huerta et al., 2016; Lee, 2015; Weaver et al., 2010), or traditional interviews and focus groups (Oh et al., 2015; Okoniewski et al., 2014). This study took a different approach, using a context driven approach to directly observe the online health information seeking behaviors of those who have been told by a healthcare professional that they have an increased risk of ovarian cancer due to family health history or a confirmed pathogenic genetic variant. During these observations, participants utilized a consistent strategy for locating and assessing online information. Starting with a Google search, participants used key words to identify likely information sources, sought out familiar resources, and assessed unfamiliar sources by learning more about the organization behind the website, gauging professionalism via the user experience, and seeking consensus. Information seeking behaviors that participants exhibited in response to the prompts provided, align with prior research findings and agreed with current National Institutes of Health (NIH) recommendations for evaluating health information on the internet (National Institutes of Health, 2011).

Participants sought out resources with which they were familiar with when reviewing search results. This observation aligns with related research, suggesting that people tend to use information sources that they are familiar with, especially during high stress times, and that people are more likely to believe and act on a message if it comes from a familiar source (Steelman et al., 2015). These findings are also supported by a 2013 Nielsen report, which indicates that individuals with internet access are more likely to return to familiar online sources (Nielsen Holdings, 2018), and a 2014 PEW research report, identifying familiarity as an

important information characteristic (Mitchell, 2014). This finding highlights the link between online information source familiarity and trust, as participants were more inclined to address their informational needs by seeking familiar sources, implying the belief that information contained on these familiar sites was more likely to meet their needs.

This finding also has implications for the role of branding. A study published in 2011 found that “brand recognition” was important to individuals navigating health information online (Powell et al., 2011). People favored online resources with “real-world” branding, defined as an identity that they recognized from their offline experiences (Powell et al., 2011). Branding that is memorable, easy to distinguish, and has a “real-world” presence may help an information resource stand out to the information seeker engaging in a quick review of search results (Eysenbach & Köhler, 2002). This concept has certainly been leveraged with general ovarian cancer awareness efforts through the use of the color teal and is widely recognized in the broader field of marketing, where brand familiarity is strongly linked with brand choice (Nielsen Holdings, 2018).

Interview participants also consistently referred to organic search results on the first page of search results. Google uses algorithms to return search results that fit the search query of the user the best, and ranks pages based on how useful they will likely be to the user (Bullock et al., 2018; Google, 2018). These algorithms assess various factors to identify the webpages that update their content frequently, is geographically close to the where the search is being conducted from (for organizations with a brick and mortar presence), that mention search terms often, and to assess the pages user experience or how user friendly it is (Google, 2018). Google also looks for websites that other users appear to value for similar online searches (Google, 2018). This method of ranking webpages may offer some level of quality control, specifically in prioritizing pages that have recently been updated, another recommendation from the NIH’s

recommendations for evaluating online health information (National Institutes of Health, 2011). This should not be the only information quality control tool in place, however, since webpages that appear in search results may enlist the help of search engine optimizations (SEO). SEO impacts the online visibility of a webpage in a search engine's results, like Google, increasing a webpages prominence within search results (Zhang & Cabage, 2017). This may include writing content with specific keywords or phrases, frequent content updates, and utilizing a webpages meta data to add underlying descriptions of the webpage that search engines may see and rank higher in search results. This is the delineation between "organic" search results, or those that rank higher as the result of prior use by other users and routine content updates, and "earned" search results, or those that have improved their ranking with SEO.

Participants who talked about website usability, mentioning font and appearance, mirrored website usability recommendation in their comments. Current usability recommendations speak specifically to font size, the use of attention-attracting features (such as images and infographics), and color-coding, or use of high contrast (Schmutz et al., 2016). This finding is not surprising, as website usability recommendations are generated through usability testing, exploring user preferences and the user experience, asking users to talk specifically about things they like or dislike about online material.

Participants' tendency towards specific TLDs, such as .org, .edu, .gov, and away from .net, also aligns with current research (Chi et al., 2018). This finding also lines up with the NIH's recommendations for how to evaluate health information on the internet (National Institutes of Health, 2011). The first recommendation from the NIH encourages individuals to identify who runs a website, and who pays for the website, citing the TLD as a starting point to acquire this information, something that interview participants appeared to enact in the searches that were observed. For example, a website that ends in with the TLD of .gov, is a U.S. government funded

website. A website that ends with the TLD .edu, is typically maintained by an educational institute. One run by a noncommercial organization typically will use the TLD .org, and finally a .com suggests that a commercial organization owns and operates the web address. By seeking this information first when reading search results, participants were successfully enlisting the first two NIH recommendations for assessing online health information (National Institutes of Health, 2011).

Unfortunately, current regulations around the use of certain TLDs suggest that the TLD may not be a reliable way to assess website quality. For example, while .org suggests that a non-profit organization is behind the online resource, there are no laws which restrict other organizations from using this TLD. Currently, .gov, .edu, and .mil are the only TLDs with restrictions in the US; .gov can only be used by divisions of the federal, state or local government .edu is restricted to public universities and other educational institutes, and .mil is restricted to divisions of the U.S. military (Gardos et al., 2015). Other TLDs, even those traditionally associated with trustworthy sources, may be purchased as part of the URL by organizations who may have ulterior motives which may impact information quality and accuracy.

Participants often visited, or looked for, the “about us” page, when assessing an unfamiliar webpage. This is the third NIH recommendation, which encourages individuals to seek documentation about the website (National Institutes of Health, 2011). Participants also talked about when the website was last updated, as an important assessment criterion, favoring resources that were updated more recently. This also aligns with one of the NIH’s recommendations for vetting online health information, encouraging users to seek sources that have been updated or reviewed recently. The NIH guidelines specifically say that even if the

information has not changed, a quality website will still be reviewed periodically to ensure that the information is all still valid (National Institutes of Health, 2011).

Interview participants were disinterested in looking up symptoms online, explaining that they did not want to worry themselves, adding that online symptom checkers tend to put the worst possible scenario front and center. A 2015 study exploring the relationship between cancer worry and information seeking found that worry is associated with cancer information acquisition for some, and avoidance behavior for others (Nelissen et al., 2015). That study surveyed a group of individuals with a cancer diagnoses and a group of individuals without cancer and found that a cancer diagnosis did not moderate an individual's information seeking behaviors. However, the study failed to examine if, or how, a genetic predisposition may impact symptom information seeking behaviors.

The opinions expressed by interview participants here suggest that for individuals with a family history or confirmed genetic predisposition of ovarian cancer, there is a difference between looking up symptoms and looking up information about a confirmed genetic predisposition. No one was interested in looking up symptoms, but most study participants said the internet was a useful resource to bridge the information gap between positive genetic test results and when their healthcare provider appointment took place. Participants said online information during this time could be used to inform their discussion with their healthcare provider. Other research supports this observation, suggesting that health care providers are still a trusted source of information, with cancer information seekers relying on their physicians to assist in interpreting what they find online (Hesse et al., 2010; Eysenbach et al., 2007).

Participants also talked about how useful it would be to have their health care provider recommend a trustworthy resource, removing the guess work from wading through a sea of

online health information. Prior research speaks to “information prescriptions” that are intended to direct people to websites with quality and accurate medical information (McKnight, 2014). These “information prescriptions” are not necessarily a response to public demand, however, but rather intended to help physicians save time that they would otherwise have to spend addressing patient questions generated from consuming poor-quality online health information.

There was one participant, however, who said during this time they would “stay off the internet.” Study participants seemed to assign a power to online health information, in looking up symptoms online or searching for information related to a genetic risk. This information would either inform, or it would increase worry. Across all interviews and scenarios, the issue of worry seemed to be more tied with symptoms, aligning with prior research on “cyberchondria”, or online health-related information seeking fueled by anxiety about one’s health, that amplifies an individual’s health anxiety (te Poel et al., 2016). Future research should explore how online information seeking impacts the health anxiety of individuals with and without a confirmed genetic predisposition.

Social media came up as a vehicle for sharing genetic test results, but only for one family represented across interviews. This aligns with what is currently known about how patients use social media to share information. Social media users are very supportive of the idea of using social networking sites to collect and share family health history (Welch et al., 2015), including genetic information (Moore et al., 2018). Patients with newly-described genetic findings reportedly turn to social media to find and connect with others (Rocha et al., 2018; Greenwood et al., 2016). While there are web-based social networking sites that specifically served as hubs of health information sharing within families (e.g. Simons Variation in Individuals Project (www.simonsvipconnect.org) and GenomeConnect (www.genomeconnect.org), there is a case for using existing webtools, like Facebook, that users are familiar with. Most social media users

are comfortable sharing medical information in private social media groups, like closed or secret Facebook groups, targeted to families alone (Rocha et al., 2018). Closed Facebook groups, like the one mentioned by a participant in this study, can be used to facilitate communication with family members, including those who are out of touch or estranged, potentially improving proband reach and information dissemination. This method of communication, however, is not without privacy concerns (Househ et al., 2018), making this an area of hereditary cancer health communication that requires more research.

This study demonstrates that individuals with an increased risk of ovarian cancer seek familiar resources and assess unfamiliar resources in a manner that aligns closely with the first three NIH recommendations for assessing online health information. Generally speaking, individuals with an increased risk of ovarian cancer appear to be comfortable using the internet to look up information about ovarian cancer but are less inclined to use the internet to look up information about ovarian cancer symptoms. Participants noted that they did not want to worry themselves by looking up ovarian cancer symptoms. The relationship between cancer worry and information avoidance has been explored in prior research, but little is known specifically about how genetic predisposition, or lack thereof, may impact information seeking. This is an area of study that warrants additional exploration.

Social media did not appear to play a major role in how people seek ovarian cancer information. But it did come up when participants talked about how they share information about genetic predispositions with their family. These results suggest that closed social media groups may serve as an important communication tool, enabling individuals to easily share the “information prescriptions” they want to receive from their health care providers, with other

family members who may be impacted by this information. More research is needed to better understand how closed social media groups may address this communication need.

Limitations

One of the primary limitations for this study was that, even though participants subscribed to a relatively consistent method for assessing online information seeking, this study did not take a look at whether or not their assessment was accurate. In other words, just because online health information seekers feel like they can discern quality information from poor quality information, or misinformation, does not mean they actually did. Furthermore, the key characteristic that participants focused on was trustworthiness of an online resource, or if it had the markers of a website they felt they could trust. Again, without further research, it is unclear whether or not trustworthiness and familiarity are characteristics that actually lead users to engage with high quality information sources in the context of ovarian cancer online health information seeking. Another limitation relates to the study sample. Since most participants were recruited via online social networks, they were likely representative of a population well versed in using the internet to find information. It would be interesting to target the more relaxed internet user, to assess their online health information seeking behaviors.

AIM 3

Introduction

Strong participant recruitment practices are critical to public health research, but are difficult to achieve (Reuter et al., 2018; Treweek et al., 2018; Briel et al., 2016). Traditional recruitment practices are often time-consuming and costly and fail to adequately target difficult to reach populations (Dew et al., 2013, McIntosh et al., 2000). Radio or newspaper advertising, referrals, or flyers all may have a wide reach, but it is difficult to target specific populations using these approaches. The initial net cast using these types of recruitment methods may yield a high number of interested parties, but such efforts result in proportionately few eligible, enrolled participants, and certain demographics are often left underrepresented (Yancey et al., 2006). Traditional recruitment efforts are also often difficult to change and adapt once they are initiated or published (Graham et al., 2008). According to the National Center for Advancing Translational Sciences (NCATS) this challenge needs to be addressed by generating “evidence-based strategies to trial participant recruitment and patient engagement,” (National Center for Advancing Translational Sciences, 2018a & 2018b).

Social media is well positioned to address this area of need, providing new platforms for potential participants to learn about public health research (Caplan et al., 2017; Topolvec-Vranic & Natarajan, 2016; Fenner et al., 2012; Temple & Brown, 2012). The term social media broadly describes online social networking platforms, where the public is able to generate, engage with, and share information, and include platforms like Facebook, Twitter, and Instagram (Obar & Wildman, 2015; Dizon et al., 2012). Social media enables researchers to deliver information to a wide audience, target specific groups of people including hard-to-reach subpopulations (Caplan et al., 2017; Topolvec-Vranic & Natarajan, 2016; Fenner et al., 2012; Temple & Brown, 2012),

and adapt outreach efforts on an ongoing basis with ease (Graham et al., 2008). Current research favors social media recruitment methods over traditional methods, with respect to both cost and efficacy (Gorman et al., 2014; Seltzer et al., 2014; Yuan et al., 2014; Rabin et al., 2013).

Facebook is particularly well suited for research recruitment. Over 69% of US adults with internet access use some type of social media, including Facebook, YouTube, Pinterest, Instagram, and Twitter (Pew Research Center 2018a & 2018b; Krogstad, 2015). Current research indicates that Facebook is an effective platform to use to recruit certain hard to reach groups, including individuals impacted by HIV and childhood cancer—demographics that are notoriously difficult to reach through traditional recruitment efforts (Seltzer et al., 2014; Yuan et al., 2014; Gorman et al., 2013; Rabin et al., 2013). With more than 207 million Facebook users from the United States in 2018, a number that is forecasted to grow to 220 million by 2022, Facebook is used by more than three-quarters of online teens and adults (Statista, 2018; Duggan et al., 2015). Facebook users engage with user generated content, posting photos to their Facebook pages, status updates, and sharing information with friends and family. Users follow content of interest, including information about various organizations and products, and socially engage with paid advertisements and other content. Compared to Twitter, which limits user generated content to 140 characters, Facebook, without any limits, invites additional interaction and discussion. Researchers can capitalize on this environment, creating paid advertisements and unpaid posts, with imagery and content tailored to catch the attention of specific populations (Chiang et al., 2016). It is this climate that has paved the way for online behavioral advertising and respondent driven sampling, where participants refer others to the study. Similar to snowball sampling, this type of sampling is helpful for reaching difficult-to-reach populations.

Facebook allows researchers to use online behavioral advertising (OBA) data to increase and improve marketing reach. OBA data is a body of information collected from a broad range of online sources about behaviors that users exhibit online (Curtis, 2014). This cache of data includes details about a user's online activity, including what websites they frequently visit, search terms they use, details about employment history, education, ethnicity, interests, even location and income (Curtis, 2014). This information is tracked and collected by third party entities that essentially store data and make it available for use through targeted advertisements. When a researcher pays for an advertisement over Facebook, or any other social media entity, they are paying for the opportunity to use this information to target a specific audience. This approach to research participant recruitment appeals to researchers in the field of public health (Curtis, 2012; Voytek et al., 2012; Bull et al., 2011; Carpenter et al., 2011; Hagan, 2010). Instead of wondering if a flyer is posted in the right place for the right type of individual to see, researchers can guarantee that their message is being displayed to the intended person. Despite the potential for targeted communication through social media, some health professionals and researchers have expressed reluctance about this innovative communication tool. These individuals are concerned about the quality of online health information (Banki et al., 2016; Sobota & Ozakinci, 2015; Kukreja et al., 2011; Friedman et al., 2011), biased sampling or reach that may accompany social media platforms (Arigo et al., 2018), and concerns about privacy (Dizon et al., 2012; Mostaghimi et al., 2011; Greysen et al., 2010).

Facebook also allows public health professionals to leverage existing social networks through snowball sampling (Baltar & Brunet, 2012; Brickman Bhutta, 2012). Snowball sampling takes advantage of existing online relationships and how individuals share information across social networks (Brickman Bhutta, 2012). By encouraging a small sample of a target population

to recommend others to a study, snowball sampling helps researchers access hidden subpopulations that are typically difficult to sample through traditional recruitment methods (Brickman Bhutta, 2012). It is this quality that grants social media recruitment the ability to potentially shift the pattern of health inequities, improving representation of certain communities in the research arena (Griffiths et al., 2012).

Facebook advertisements also invite use of tone and imagery, which may facilitate research recruitment. Encouraging participants to recommend the study to friends and family does more than just bring new people into the study. It also instills a sense of altruism for the referrer, giving them the impression that they are passing on information that may be beneficial to a friend or family member. Research has shown that women respond more favorably to preventative health care and associated research when they believe their participation or action will be of some benefit to their family or friends (Newington et al., 2014). Tone and imagery may modify this effect (Rus & Cameron, 2016).

Finally, Facebook, and other social media platforms, invite researchers to leverage networks around patient advocacy groups (MESM, 2017). Many patient advocacy groups boast an active social media presence with an established social network. By networking with these groups, researchers can work to ensure that their followers see recruitment material, using these online communities as a launching point for snowball sampling and tailoring advertisement material with these audiences in mind. From snowball sampling, to inviting opportunities to shape tone, imagery, and content to fit the needs of the intended audience, social media is well positioned to function as a targeted communication tool.

Recent reviews indicate that most studies utilizing Facebook to recruit participants for health research focused on people who were 18 to 30 years of age (Whitaker et al., 2017;

Topolovec-Vranic & Natarajan, 2016). In comparison, few studies have evaluated social media as a means of recruiting people impacted by cancer who are 35 years of age and older (Whitaker et al., 2017), and no studies have explored how social media recruitment performs when targeting women at risk for ovarian cancer. The consensus is that older people may be less likely to adopt new technology, like social media (Whitaker et al., 2017; Kapp et al., 2013). For example, Kapp et al. (2013) found that women between the ages of 35 and 49 were difficult to recruit via Facebook for health research participation. Other studies have reported high reach, but low engagement among social media users, resulting in a high attrition rate for social media recruitment (Bajardi et al., 2014). But studies of this nature fail to examine ad content or take the growth of the social media platform into consideration. Targeted advertisements may improve attrition rates, using OBA combined with content that speaks to the intended audience. Additionally, as the social media base continues to grow, the profile of the average user evolves, and with it the age of the average Facebook user continues to increase (Frandsen et al., 2014). With this evolution in mind, ongoing assessment is needed to evaluate the efficacy of social media for research participant recruitment across different demographics and more research is needed to better understand how Facebook functions as a tool for recruitment in the context of ovarian cancer (Curtis, 2014).

Study Aims

This study sought to determine if Facebook is an effective recruitment tool for targeting women 30 years of age and older for recruitment into the MAGENTA (Making Genetic Testing Accessible) Study by evaluating innovative methods for the recruitment of research participants using Facebook. To accomplish this objective this study generated a series of posts and advertisements, utilized a variety of imagery and language, and leveraged Facebook's OBA tools

to target specific populations and eligible participants. Paid advertisements and unpaid Facebook posts were posted to a diverse selection of Facebook pages, as well as groups with a demonstrated interest in ovarian cancer research, including patient advocacy groups. Next, quantitative data reporting patterns of responses, including engagement, response rate, and enrollment were collected from all Facebook posts and advertisements and analyzed on an ongoing basis, for each type of posting or advertisement. We hypothesized that unpaid Facebook posts and Facebook advertisements will improve reach of study material, and result in improved study enrollment. Results from this study can be used to determine if Facebook is an effective platform for recruiting women who are 30 years of age or older, for an ovarian cancer research study.

Methods

About the MAGENTA Study. The MAGENTA (Making Genetic Testing Accessible) Study (<https://magenta.mdanderson.org/>) is a nationwide Stand Up To Cancer (SU2C) initiative, seeking to improve access to genetic testing for ovarian cancer, recruiting and randomizing 3000 women, from across the United States, with an increased risk for ovarian cancer due to family or personal health history. MAGENTA participant recruitment took place between April 2017 and October 2018, in-person and online, in collaboration with efforts from patient advocates around the country. Participating patient advocacy groups included the Minnesota Ovarian Cancer Alliance (MOCA), the National Ovarian Cancer Coalition (NOCC), and the Ovarian Cancer Research Fund Alliance (OCRFA). Once potential participants learned about the MAGENTA study, they were prompted to visit the study website, which discussed eligibility criteria and provided additional information about the MAGENTA study. From there, interested parties clicked to participate in the subsequent online communication system, from study consent to study questionnaires. The online communication system is illustrated in Figure 7.

Identifying study collaborators. Before designing any study materials, the first step was to identify key study collaborators to assist in generating and disseminating study materials, in person and over social media. This group consisted of a number of study collaborators, including health care professionals from cancer care and research centers and patient advocates from advocacy groups across the US. Cancer research centers included Stand Up To Cancer (SU2C), MD Anderson Cancer Center, and American Association for Cancer Research (AACR). MAGENTA patient advocates included SHARE Cancer Support (<https://www.sharecancersupport.org/>), Ovarian Cancer Research Fund Alliance (OCRFA) (<https://ocrahope.org/>), National Ovarian Cancer Coalition (NOCC) (<http://ovarian.org/>), and the

Minnesota Ovarian Cancer Alliance (MOCA) (<https://mnovarian.org/>). Both cancer research and treatments centers, as well as the patient advocacy groups were all members of the Stand Up To Cancer-Ovarian Cancer Research Fund Alliance-National Ovarian Cancer Coalition Dream Team. Patient advocates were extensively consulted during the development of the study materials, including the media kit described below.

Developing a media kit. Adapting methods outlined in Carter-Harris et al. (2016) and Musiat et al. (2016), the first step in this study was to develop a media kit, including Facebook recruitment materials. The media kit included a list of social media contacts, such as patient advocacy groups and other groups with an apparent interest in breast and ovarian cancer. Breast cancer organizations were included because a personal or family history of breast cancer was one of the eligibility criteria for participation in the MAGENTA study. Colorectal cancer organizations were included because of the close genetic relationship between colorectal cancer and ovarian cancer (Soyano et al., 2018). This step involved completing an inventory of Facebook pages with an ovarian cancer connection, by searching for “ovarian cancer,” “breast cancer,” and “colorectal cancer” via Facebook’s search engine. The top 30 Facebook pages and groups identified through this means were explored, with the goal of identifying related pages and groups, via snowball sampling.

Search results over Facebook’s search engine are returned in order of relevance (to the search terms used and user conducting the search) and social media activity, favoring pages and groups with recent activity and a high number of followers or members (Wable, 2010). Patient advocacy groups identified through this process were contacted for related referrals by email or by phone. Numbers of followers were noted for each of selected Facebook pages and groups and were ranked by activity favoring pages and groups that were more active with a greater number

of followers. A contact person or persons was identified for each Facebook page or group, if available. This person served as the primary “contact person” and was approached about posting an unpaid ad on behalf of the study. When no contact information was available, groups and pages were contacted directly over Facebook messenger with information about the study.

In addition to this directory of contacts, the media kit also included different types of posts generated for recruitment purposes, including paid advertisements (Figure 8) and unpaid posts (Figure 9), sample tweets, a list of relevant hashtags to incorporate into posts, and a selection of media for use across all social media posts and advertisements. A MAGENTA Facebook page was also generated, in an effort to develop trust with potential participants (O’Connor et al., 2014; Walker, 2013). The Facebook page provided basic information about the study, served as a platform for sharing unpaid and paid social media posts, and directed potential participants to the study website. Materials from the media kit were assessed by study patient advocates and Facebook posts underwent usability testing, prior to IRB approval.

Advertisements and posts were created with tone and imagery in mind, focusing on content related to ovarian cancer research that elicits a combination of the following concepts, adapted from Batterman (2014):

- 1) Content instills a sense of collaboration, conveying the idea that one is participating in research as a member of a team, in order to address a health problem (in this case, ovarian cancer was framed as the problem).
- 2) Content instills a sense of independence or conveys the idea that one is addressing the problem of ovarian cancer as an individual through research participation.
- 3) Content instills a sense of altruism, conveying the idea that the individual is participating in research for the benefit of others.

- 4) Content instills a sense of self-gain or self-preservation, conveying the idea that the individual is participating in research for personal gain.

A total of 16 core paid advertisements and unpaid posts were generated for use over Facebook. Unpaid Facebook posts and paid advertisements included at least one media component, such as an image or the study video, as well as a brief description of the study, and any relevant hashtags. Various text and media components were intended to be rearranged, to generate different posts, inviting easy modifications. Each paid advertisement and unpaid post also included a link to the study homepage, where potential participants were encouraged to review study criteria and begin the enrollment protocol from the MAGENTA study, if desired (Figure 7).

All advertisements were produced into different formats, to accommodate different device types, including the following: 1) desktop news feed (capturing individuals who find the study from their desktop computers); 2) mobile news feed (capturing individuals who find the study from their handheld mobile devices); and 3) desktop right column. Facebook offers varied formatting to accommodate different device types, so this list is not exhaustive and is updated regularly by online social media platforms. These variations reveal what devices potential study participants are most likely to use to interact with a paid Facebook advertisement. All paid advertisements and unpaid posts were reviewed, and adapted as needed, prior to implementation, by study personnel and patient advocates connected to the MAGENTA ovarian cancer study.

Publishing unpaid posts. Unpaid posts were published directly to the MAGENTA Facebook page. In addition to publishing study material to the study Facebook page, a member of the research team also contacted a designated page administrator or patient advocate in charge of a Facebook page or group identified in the initial Facebook page inventory. These individuals

were contacted via phone or email and asked to publish a series of unpaid posts to their Facebook group or page. When it was not possible to contact page administrators by phone or email, they were contacted via Facebook messenger. Individuals who did not respond within a week received a follow up message. Once contacted, page administrators and patient advocates were asked to post the unpaid study notification to their page on behalf of MAGENTA. Early in the recruitment process, no more than five ovarian cancer related Facebook groups were prompted to publish study material at any one time to help moderate enrollment and assess the efficacy of the online communication protocol. For example, by limiting the number of posts published at any one time, posts could be staggered or halted all together in response to a dramatic increase in website activity or enrollment, if needed.

Publishing paid advertisements. Paid advertisements were generated through Facebooks advertising tool. This tool prompts the user to select a marketing objective. Possible objectives include awareness, consideration and conversions. The “awareness” objective includes “post promotion.” “Post promotion” is an option on Facebook that allows users to “boost” a specific post or pay to ensure that more users see that post. The “consideration” objective includes “traffic generation.” “Traffic generation” allows users to pay Facebook to drive traffic, or website visitors, to an external URL like the MAGENTA study homepage (Figure 7). The “conversions” objective allows users to pay Facebook to drive traffic to an external page webpage and make sure individuals who land on the webpage view page material. These objectives are also defined in Table 9. Paid advertisements generated on behalf of this study included “post promotion,” or “reach,” “traffic” to the MAGENTA homepage, and “conversions” to the MAGENTA homepage.

After the marketing objective was identified, the tool prompts the user to define their preferred audience using Facebook's audience targeting tool. Facebook allows users to generate a targeted audience based on geographic location, language, age, gender, as well as other demographics, interests and behaviors. Targeted populations for the purposes of this study included English-speaking women over the age of 30, living in the United States. Additional geographic and behavioral targeting were included on a case-by-case basis and are described in greater detail below (Table 10). Census maps were used to select additional geographic and socio-economic targeting, including expressed interest in a breast or ovarian cancer organization, proximity to a minority-majority location, or rurality of the location. These variables were layered using ArcGIS to select specific geographic targets. ArcGIS is a mapping and analysis tool that allows users to utilize a geographic information system (GIS) to capture, manipulate and analyze spatial and geographic data.

Facebook also allows users to create custom audiences. In creating a custom audience, the user chooses an audience source (e.g. a specific Facebook page or an external URL, like <https://magenta.mdanderson.org/>), and then Facebook will identify common qualities of people who visit that online page, using a Facebook pixel. The Facebook pixel is an analytic tool that is triggered when someone visits the website it is loaded on to. The pixel subsequently collects publicly available information about website visitors and the actions they take on the website and can be used to inform and assess advertising efforts (Facebook, 2018A). Once demographic information and user interests are selected, manually or informed by pixel data, Facebook finds people who "look like" these users (Facebook, 2018A). This is called a Lookalike Audience. Targeted audiences are summarized in Table 10.

Advertisements were placed using the “automatic placement” tool. This option helped to maximize the budget and ensured that Facebook’s advertisement delivery system allocated the advertisement set’s budget across multiple placements, based on where they are likely to perform best. Advertising content was then uploaded. Content included text, the URL for the MAGENTA study homepage, and a media component. The media component was either the study video or an image. Finally, ad budget and schedule were selected. This included selecting either a daily budget or a lifetime campaign budget and identifying how long the advertisement would be running.

Other recruitment efforts. All Facebook recruitment efforts took place alongside other traditional recruitment efforts as part of the MAGENTA study. Traditional recruitment efforts included clinician referral, email, disseminating study information at provider and patient advocate conferences, and disseminating flyers in patient advocacy and clinical settings. Traditional efforts were based out of participating cancer research institutes, organizations, and patient advocacy groups. For example, OCRFA disseminated flyers at cycling fundraising and awareness events that they host at different locations across the US. All recruitment efforts took place within the US.

Evaluating paid advertisements and unpaid posts. Facebook analytics were used to capture how users responded to MAGENTA social media content, tracking unpaid posts published on the MAGENTA Facebook page, as well as paid advertisements. Facebook analytics included “engagement,” “results,” “reach,” “impressions,” “clicks,” and “video plays,” among others. All analytics and their method of calculation are summarized in Table 9. Facebook defined “engagement” as the number of social media actions users took after viewing an ad or post. In other words, instances when they saw an ad and clicked on it or engaged in some other

manner. These actions may include instances when a user “likes” a post, “comments” on it, or “shares” it across their social media page. “Clicks” is defined as the number of users who click on a link associated with an ad or post. “Video plays” is defined as the number of times a video posted to Facebook starts to play. “Impressions” is defined as the number of times an advertisement was on a screen. “Reach” is defined as the number of people who saw content. “Results” is defined as the number of times the advertisement achieves an outcome, based upon the campaign objective. “Engagement,” “clicks,” and “video plays” can be observed across paid advertisements and unpaid posts, while “impressions” and “results” are only reported on paid advertisements (Facebook, 2018B).

Facebook analytics were tracked on a daily basis with the release of any paid advertisements. During each paid campaign, and over the course of the two weeks following each paid advertising campaign, enrollment was tracked, with a focus on any relevant geographic information, including regions within 25 to 50 miles of cities with a minority-majority. When a paid advertisement was not in progress, analytics were managed on a weekly basis. Study enrollment was tracked on a weekly basis. Page administrators, who posted MAGENTA material on behalf of the study, were asked to forward analytics on post-performance, if available, and to update study personnel when they published MAGENTA posts. MAGENTA study participants were also asked to report how they heard about the study in a required eligibility questionnaire, informing how many potential participants learned about the study through clinician referral, email, flyers, or social media.

Analytics for each advertisement campaign were reviewed daily to assess campaign performance. For example, was the advertisement reaching the intended target and was it generating recruitment activity (e.g. like clicks to the MAGENTA homepage)? Daily review

provided opportunities to make adjustments to the campaign content or targeting. The same information was collected for unpaid posts that were published directly on the MAGENTA Facebook page. Analytics from other unpaid posts published to other Facebook pages or groups, were retrieved from the administrator responsible for the post on a monthly basis, when available. If at any time the MAGENTA study website or another part of the online communication system became overburdened, for example when there was too much web traffic for the MD Anderson MAGENTA homepage to handle, advertisements or unpaid posts were pulled, or turned off, until traffic subsided.

We also observed the Facebook specific costs associated with paid advertisement campaigns. Cost was measured based on the total spent over the lifetime of the campaign as well as “cost per result.” “Cost per result” was a measurement of how much each result, or outcome, cost, where results may include number of website views, Facebook page engagement, or reach. For example, posts that were intended to get potentially interested parties to watch a study video, delivered results for the lifetime cost of the individual campaign, but also provided data about cost per result, or in this case, “cost per video view.” Eligible enrollments resulting from participant-reported online recruitment efforts were assessed, with a focus on reach, interested parties, and successful enrollments.

Results

MAGENTA study recruitment took place between April 2017 and October 2018. The recruitment timeline can be viewed in Figure 10. Social media recruitment started September 2017 and continued until study recruitment goals were met. Prior to September 2017, the MAGENTA study relied on traditional recruitment methods. Over the course of recruitment, Facebook materials including paid advertisements and unpaid posts, had a reach of 407,769, resulting in 57,248 instances of engagement, suggesting that around 14% of people who saw information about the MAGENTA study over Facebook engaged with the content. Reach and engagement excluded posts that were published on Facebook groups and pages managed by other breast and ovarian cancer related groups. During this time the MAGENTA study homepage was shared 1,948 times, and the MAGENTA study video was viewed 31,358 times (Table 12).

Users learn about the MAGENTA study over TV and social media

Of the total 13,983 who responded to the MAGENTA REDCap eligibility questionnaire, less than 1% (n=23) indicated they learned about the study from a magazine, just under 1% (n=86) indicated they learned about the study from the radio, just under 2% (n=253) said they learned about the study from a healthcare provider, 3% (n=459) said they learned about the study from a patient advocacy group, just under 8% (n=1102) from a friend, around 9% (n=1209) from a family member, whereas 27% (n=3837) indicated that they learned about the study online, either from social media or another webpage, and 28% (n=3938) said they learned about the MAGENTA study from the television. Among those who reported they learned about the study from the internet, 17% (n=2369) specifically cited social media. The remaining 18% of individuals who responded to the REDCap eligibility questionnaire did not indicate where they heard about the study from.

Social media response rate

A breakdown of responses by paid advertisement campaign can be reviewed in Table 11. Paid advertisements had a total reach, defined in Table 9, of 373,682. Among those reached, about 4% clicked on a link (n=13,357) and just under 15% engaged with page content (n=54,117). Paid campaigns also generated 19,792 video plays and 9,095 website conversions, defined as instances where a potential participant viewed page content on the MAGENTA homepage. Details about the homepage can be reviewed in Figure 7. Paid advertisements utilizing the study video resulted in a total reach of 54,992 and 28,586 instances of page engagement. Post promotions, or paid advertisements that focused on increasing reach of a post, resulted in 2,666 reach and 97 engagement. Conversion campaigns, focusing on getting potential participants over to the MAGENTA homepage to view content, resulted in 268,052 reach, 35,904 engagement and 9,095 conversions. Campaigns seeking to drive traffic to website, different from conversion campaigns which aim to get users to visit and view page content, resulted in a reach of 80,120, 18,158 engagement, and 1,697 instances where a unique user clicked on the link to the MAGENTA homepage.

Almost all users engaged with paid advertisements from a handheld mobile device, like a smartphone or tablet, rather than via a desktop computer. Most users, just over 66%, engaged with paid advertisements from their Android device (n=35,806), and just over 30% engaged from an IOS devices (iPad, etc.) (n=16,466). The remainder engaged in material from either a Desktop computer or another type of device. Around 9% of women younger than 54 years of age reached by advertisement content engaged with the ad, while almost 26% of women between the ages of 55 and 64 who saw the paid content engaged, and finally, 44% of women 65 year of age or older who saw MAGENTA ads engaged with ad content. Unpaid posts published to the MAGENTA

Facebook page resulted in a total of 34,087 reach and 3,131 engagement. These numbers do not include social media posts published to other, non-MAGENTA Facebook groups and page.

Facebook currently does not have a way to track website conversions through unpaid posts, or a public facing means of tracking demographics of engaged users or platforms they access content from, so this data point was not collected for unpaid published materials.

MAGENTA study enrollment and randomization summary

The MD Anderson MAGENTA homepage received a total of 34,715 unique visitors and 22,029 unique clicks during the enrollment period. Over the course of the campaign, about 64% of users who visited the MD Anderson MAGENTA homepage clicked on the “Get Started” link, directing them over to the landing page on the REDCap system. The “Submit” button on the REDCap landing page received a total of 14,025 clicks, and the eligibility questionnaire on REDCap was complete 10,883 times. Of the completed questionnaires, 14.1% (n=4,887) were eligible. A total of 3,422 eligible individuals consented to participate, or 9.9% of the users who originally visited the MD Anderson MAGENTA homepage. A total of 2,851, or 8.2% of users who visited the MD Anderson MAGENTA homepage, completed the baseline questionnaires on REDCap and were randomized. Enrollment and randomization data are summarized in Table 12.

Social media campaigns and news stories influence enrollment response

General recruitment activity following paid advertisements was tracked and compared to time periods when paid advertisements were not running. Due to overlap in campaigns and television news stories, changes in recruitment activity around paid campaigns is not reported for all campaigns, and in some cases the observation period following the campaign was excluded due to another campaign running during that time. Completed eligibility questionnaires were referenced as a more immediately observable sign of impact, as opposed to study consent, since

it took participants some time to move all the way through the online communication system. For example, a participant may complete the eligibility questionnaire the same day they saw the advertisement but may not consent to participate until later. There was an uptake in completed eligibility questionnaires following individual and successive paid campaigns. Prior to two paid advertisements, which ran back to back in November 2017, there was an average of 5.2 eligibility questionnaires completed daily. This number increased to 6.9 during these campaigns and in the week following the campaign. During the two weeks prior to another pair of paid advertisements, published sequentially in March 2018, there was an average of 7.4 completed eligibility questionnaires daily, increasing to 12.7 during, and in the two weeks following the campaign.

Enrollment following paid advertisement campaigns with a narrow geographic focus was further assessed. These campaigns included a targeted campaign in Washington State (WA Campaign) and a campaign with multiple advertisements in California (CA Campaign 1, CA Campaign 2, and CA Campaign 3). A detailed overview of these campaigns is provided in Table 10. Results from the campaigns are summarized in Table 11. The MAGENTA study briefly ran a targeted campaign in New York as well, but this campaign was excluded from enrollment response analysis since enrollment was not open in New York prior to this campaign. The WA Campaign, reached over 20,733 people, about 1% (n=298) of whom clicked on the webpage link, and 73 went on to view content on the MD Anderson MAGENTA page. During the campaign a total of 32 individuals from Washington State completed the eligibility questionnaire, for a rate of 3.7 completed eligibility questionnaires per day. Potential participants completed eligibility questionnaires as a requirement for enrollment (the progression of these steps can be reviewed in

Figure 7). Prior to the social media campaign, there was an average of .5 completed eligibility questionnaires daily.

The ad campaign targeting California consisted of three ads (CA Campaign 1, CA Campaign 2, and CA Campaign 3). This campaign had a reach of 95,600. Just over 7% of these reaches resulted in a webpage link click (n=6,806), and 5,628 went on to view content on the MD Anderson MAGENTA page. During, and following the immediate campaign, a total of 74 individuals from California completed the eligibility questionnaire, for a rate of 1.5 completed eligibility questionnaires per day. Prior to this campaign there was an average of .6 completed eligibility questionnaires from the state of California.

During the social media enrollment period, several television news stories about the MAGENTA study were broadcast, including a story from WCCO based out of Minnesota, (<https://minnesota.cbslocal.com/2018/04/30/ovarian-cancer-megenta-test-national-trial/>), Fox 2 Detroit story out of Michigan (<http://www.fox2detroit.com/news/local-news/magenta-free-genetic-testing-helps-woman-avoid-cancer>), and the King 5 story, based out of Washington State (<https://www.king5.com/article/news/health/magenta-genetic-test-can-screen-for-ovarian-cancer/281-586362861>). Each of these stories included a video component along with a news story. These news stories were spearheaded by patient advocates and clinicians working in conjunction with the MAGENTA study and were shared widely over social media. In the month following the WCCO story, completed eligibility questionnaires from Minnesota increased from less than .5 per day, to almost 123 per day. An increase in completed questionnaires was also observed following the release of the Fox 2 Detroit story. In the month immediately following this story, completed eligibility questionnaires increased from .3 per day to 31 per day. Likewise,

in the month following the King 5 story, completed eligibility questionnaires from Washington state increased from .6 per day to 25.

Over the course of recruitment, there were several instances when recruitment trends increased, including news stories, illustrated in Figure 10. The first recruitment burst took place in September 2017, when social media recruitment first opened. During this time, a number of patient advocacy groups posted MAGENTA materials on behalf of the study in celebration of Ovarian Cancer Awareness Month. At this time the MAGENTA Facebook page also went live, publishing a variety of unpaid posts. Another enrollment burst took place following the WCCO story, published at the beginning of May 2018. Active recruitment was paused during this time in response, to help reduce the burden to the system. This was the only time social media recruitment efforts were paused in response to an increase in enrollment activity.

Discussion

This study demonstrated that Facebook is a useful way to reach women over the age of 30, who have an increased risk for ovarian cancer, through paid advertising, unpaid social media posts, and by promoting news stories over social media. For example, during and following each narrow geographic targeted advertising campaign there was an uptake in completed eligibility questionnaires from that area, and a general increase in recruitment activity during and following paid advertisement campaigns. This study also demonstrated that working with patient advocacy groups and using social media to boost patient advocate-spearheaded recruitment efforts and subsequent enrollment, is critical. The patient advocacy groups supporting the MAGENTA study were critical to the success of the study. Not only did they help to facilitate news stories, discussed in further detail below, but they also helped disseminate study information across their established online, and in-person, social networks, increasing message reach. Where visibility to anything posted to the MAGENTA Facebook page may be restricted to the MAGENTA pages' 100 or so followers, content published on NOCC's Facebook page may be easily visible to their 36,000+ followers. Relationships with patient advocates also helped to shape targeted advertisement campaigns. Using Facebook's campaign targeting tools, we were able to target individuals who followed our patient advocates Facebook pages—presumably people that may be interested in an ovarian cancer genetic testing study—as well as people who “looked” like those followers.

Following the release of various news stories, there was an increase in enrollment trends. Patient advocates were instrumental in getting these stories published and supporting story circulation. This hard work paid off, with 28% of potential participants reporting that they learned about the study over the television, referencing specific news stations featuring news

clips about the study. These stories were shared over social media, widening the reach of these efforts. Over social media more people were able to view and share the news stories than they would have been able to do without an online social network, expanding the reach and making these news features more accessible. The inclusion of visual content, like videos, in news stories appeared to extend this reach farther, making online content easier to view and share. The advantage that video media content lends to social media content is well documented, with other research confirming that video, and other media rich social media posts, perform better when compared to text-based content alone (Ordenes et al., 2018; Theiss et al., 2016).

The WCCO story was arguably the most successful individual recruitment effort. The WCCO story featured a news anchor with a family history of breast and ovarian cancer name Kim Johnson. Kim Johnson is a familiar news anchor in the Minnesota area, and had brought her personal health history into her reporting prior to the release of the MAGENTA WCCO story (for example: <https://minnesota.cbslocal.com/video/category/news-local-news/3846447-go-black-white-teal-to-fight-ovarian-cancer/>). Johnson was an established household name who had spoken about ovarian cancer in the past. It is possible that this story gained the traction it did, for the same reasons online information seekers are more likely to utilize familiar sources—if they can recognize the name, they are more likely to trust it (Steelman et al., 2015). Comparing these efforts to the enrollment activity following paid advertisements, it appears that while paid ads have an impact, collaboration with patient advocacy groups is also important for reaching a target audience. By leveraging existing social networks over social media, via patient advocacy groups, Facebook could offer more cost saving opportunities for research recruitment. In considering these opportunities, as the average Facebook user continues to age (Allen, 2013),

Facebook is likely to increasingly become a favorable venue for recruiting adults for research. A similar evolution in the average user is observable across other social media platforms as well.

News stories positively impacted enrollment trends but were difficult to modify once broadcast. For example, the King 5 story, broadcast August 2018, was originally created earlier in year. By the time it went live, some of the data contained in the story was outdated. For example, the King 5 story reported that 2400 additional participants were still needed for the study, which was no longer the case by August 2018. This was not an issue for paid Facebook advertisements, which were easy to pause and modify if needed.

Online behavioral advertising (OBA) made it easy to target information about the study to specific age groups and regions. It also made it easy to target people based on expressed and inferred online interests and preferences. For example, we were able to target people who met age and region criteria, and who had expressed an interest in various ovarian and breast cancer related initiatives. Women who were 65 years and older had the best response rate when compared to other age groups, with nearly 44% of reaches translating to engagement. This response rate suggests that even though individuals 65 and older make up a smaller percentage of online social media users, they are arguably more responsive to content they see over social media than younger demographics. For the context of the MAGENTA study, while women 65 years of age and older are eligible to participant, they are perhaps not the primary target for a study like this. Their response rate could still potentially be leveraged, but perhaps with a different recruitment message. Rather than encouraging these women to enroll themselves, future advertisements might implore them to encourage their family members to learn more about the MAGENTA study. Given their response rate, there were several campaigns that actually

included an age cut off in the advertisement targeting, excluding women who were 65 years and older (see Table 10 for age details).

Over the course of the study, the Facebook advertising service changed substantially, in response to concerns about Facebook privacy breaches in the context of OBA (The New York Times, 2018). This limited the degree to which individuals could be targeted based on demographics and required additional review of advertising content. Facebook also eliminated Race and Ethnicity as targetable demographics to prevent discriminatory marketing practices (Terdiman, 2017). Instead of targeting potential participants based on race and ethnicity, advertising campaigns sought to stratify by race/ethnicity, using current census data to identify regions with a minority-majority and targeting those geographic regions.

Facebook and other social media platforms certainly present a number of opportunities for researchers, but privacy concerns and worries over the use of OBA data make it clear that the drivers of Facebook do not always share the same values of the drivers of research. Paid advertising present unique opportunities to target specific groups of people, but unpaid posts, published across existing online social networks, are arguably preferable from an ethics standpoint, particularly in regard to recent data breaches over Facebook and concerns about how social media platforms like Facebook use and monetize OBA data (The New York Times, 2018). Data privacy issues like these impact consumer trust and may send users away from previously trusted social media platforms, like Facebook.

Despite efforts to target underrepresented communities, the majority of MAGENTA participants were white-identifying individuals. This may have been in part due to the geographic locations that recruitment bursts originated from, for example the Minnesota burst, which increased enrollment from a region composed of over 80% non-Hispanic White individuals.

Facebook claims to have a diverse user base (Chang et al., 2010), but prior work suggests that different groups have different response rates where research is concerned, meaning that targeted marketing, even over social media, is likely to leave certain groups underrepresented (Whitaker et al., 2017). Current recommendations highlight the importance of letting the target population inform the platform choice (Arigo et al., 2018). It would be worthwhile to explore other social media platforms with a minority majority, for additional minority recruitment opportunities. For example, the average Twitter user is a non-white, college educated male between the ages of 25-34, suggesting that Twitter may be a preferred social media platform for sharing information with this demographic (Twitter, 2018C). Future research should assess the efficacy of targeted minority recruitment across varying social media platforms, exploring alternative delivery models to improve access to genetic testing for women of color.

There was a fairly sizeable drop off observed from initial engagement to enrollment and randomization (Table 12). The drop off may be explained by normal study drop off at each stage, but it may also be the result of the complex online enrollment protocol used. Participants who learned about the study were referred to the study webpage, and from there were several clicks away from the eligibility questionnaire (Figure 7). Eligible individuals then had to note the messaging at the end of the questionnaire that told them to check their email inbox for an email containing the next steps and ensure that any auto-filtering system they had turned on in their email inbox did not filter the REDCap email directly in to their trash or spam box. This was an issue the came up in initial system usability testing, which was resolved by adding additional messaging at the end of the questionnaire that prompted people to check their email inbox. It is possible that some of the drop off rate between completion of questionnaire and signed consent may be due to lost emails.

Limitations

There were several limitations to this study. One of the most prominent issues was the varying definitions of “reach” and “engagement” across different online platforms. Facebook, for example, offers a different variable for unique engagement and total engagement. Total engagement encompasses the total number of actions taken. In other words, the same person could click on the same thing 40 times, and each of those individual clicks will be treated as a separate action, potentially inflating the actual engagement. Unique engagement, on the other hand, identifies that those 40 clicks were completed by the same person, and would only count one click. While Facebook differentiated these variables, REDCap did not, making it difficult to compare numbers accurately across the varying platforms included in the online communication system. This also made it difficult to determine whether or not a particular effort was very successful, or whether the participant just finally took action after seeing information about the study for the third or fourth time, a potential trend supported by marketing research that indicates that repeat exposure is required for action (Tripp et al., 1994). Similarly, Facebook does not currently have a way to track website conversions through unpaid posts, or a public facing means of tracking demographics of engaged users or platforms they access content from, so this information was not collected for unpaid published materials.

Another limitation involved the REDCap eligibility questionnaire and how it asked potential participants about how they learned about the study. Most of the individuals who completed the eligibility questionnaire said they learned about the study from an online resource or the television. Patient advocacy groups were also referenced often. It is important to note that most reported patient advocacy outreach took place online, either over e-newsletters or over their social media platforms. With this overlap, it is difficult to assess the results from this question

alone. In other words, individuals who reported they learned about the study from NOCC, for example, may have seen a post shared to their social media page, on an e-newsletter, or over NOCC's home page. Likewise, the group of individuals who learned about the MAGENTA study over the Internet, may have included folks who specifically saw information over social media. We also do not know how many times, and across how many media, participants encountered information about the MAGENTA study before they acted, or which source stuck with them as the point of origin, be it a paid Facebook advertisement or an unpaid post. For example, an individual who reported learning about the MAGENTA study via NOCC, may have previously seen an ad about the MAGENTA study over Facebook, received an e-newsletter talking about the MAGENTA study from NOCC, and saw a flyer posted in their health care providers office, before they finally clicked on a post on NOCC's Facebook page. With significant overlap between how responses to this question were defined and understood by participants, it is difficult to assess the efficacy of individual outreach efforts. Prior research indicates that posting to a variety of mediums, improves visibility, and makes it more likely that people have repeat opportunities to act over time, but additional research needs to be conducted to understand this rate of response to action functions over the internet and over social media.

Organic, or unpaid, social media response rate was also difficult to accurately capture, since there was not a reliable way to capture reach and engagement resulting from other organizations posting on behalf of MAGENTA. While several patient advocacy groups provided reports, without being able to monitor these numbers directly, along with the fact that other organizations ended up sharing study information without prompting, we were unable to capture what amounted to much of the recruitment chain-referral sampling that occurred. Future research

should utilize Facebook pixels, or a similar tracking means, to better understand the chain-referral sampling that occurs over Facebook.

Another potential limitation was the exclusion of social media user comments. Social media users did comment on the various MAGENTA paid advertisements and unpaid posts. These comments were not included in analysis, as they were not a primary source of data, but may have informed advertisement content and targeting. For example, one commenter remarked about how “creepy” it was that they had just visited a website about genetic testing, and now Facebook was showing them an ad related to this topic. This information could have potentially been used to modify targeting, by excluding the use of certain search terms in targeting.

The platform itself also presented a limitation. Platform selection appears to be one of the more important factors in conducting social media research. The most popular social media platforms currently used for research recruitment include Facebook, Twitter and Instagram. Each of these platforms has different user demographic profiles, with social media preferences varying by race, ethnicity and age. Facebook is increasingly becoming a platform that is appropriate for reaching middle-aged and older adults, as the average Facebook user ages (Bosker, 2011). Instagram and Twitter, on the other hand, may be better options for reaching younger populations, given the average Instagram and Twitter user is under 35 years of age (Instagram, 2018; Twitter, 2018C). The average Twitter user, for example, is a young, affluent, college educated non-white male (Twitter, 2018C). Certain racial-ethnic groups also tend towards other preferred social media platforms. For example, the most popular social media platform across Koreans is called KakaoTalk (Jin & Yoon, 2016). It is important to choose a social media platform populated by members of the intended population. This requires an understanding of the social media habits of the target population. It is also critical to understand that any social media

platform will be subject to sampling bias, if used to recruit research participants. Not only will recruitment activities be subject to the bias present on the specific platform, but it will also be subject to the bias that results from internet-based recruitment efforts—that is, the resulting study population will largely be made up of individuals who use the internet, a potential marker of e-health literacy and technology-literacy. Whatever the research goals, the target population should inform social media platform choice.

Conclusion

Overview of findings

This research helped to identify key methods and components that improve social media message outreach for public health messaging and research recruitment, investigating how Twitter users mention ovarian cancer, observing the online information seeking habits of individuals with a risk for ovarian cancer, and by assessing Facebook as a research recruitment platform.. Results from Aim 1 and Aim 2 confirmed the importance of integrating candid images and @mentions in social media postings. Aim 1 also revealed that social bots are high-frequency content generators that incorporate health themes into their posts. Data gathered also supported the importance of familiarity, or brand recognition. Participants described on what elements of an online information source made it more or less trustworthy, but brand recognition or familiarity was a consistent quality needed to build trust. Familiarity was important to online information seekers in Aim 2 and potential research participants in Aim 3. Overall, this research validated the importance of knowing the target audience, when designing social media materials and selecting a social media platform.

Aim 1 demonstrated that Twitter content generated by humans and social bots incorporates key structural components, including URLs, @mentions, images and hashtags. The majority of tweets collected during the data collection period were re-tweets. Accounts with a higher online follower count appeared to receive a greater number of retweets. Awareness raising events were accompanied by an influx of tweets, and social bots, which were difficult to detect, generated more tweets with a higher proportion of fear-based language when compared to human-generated tweets.

Aim 2 interview data demonstrated that familiarity is a primary criterion used in information gathering by individuals with an increased risk of ovarian cancer due to a genetic predisposition or a family history. Participants specifically sought out resources that were familiar due to prior experience, geographic proximity, or brand recognition. They also called attention to candid imagery, leaning away from “stock images,” and sought out webpages with a user-friendly interface. Participants also engaged in a consistent vetting process when they came across an unfamiliar resource. This assessment aligned with several of the NIH recommendations for assessing online health information. Overall participants did not rely on social media for information acquisition when responding to study scenarios, but one individual did mention using a closed social media group to disclose genetic testing results with family.

Finally, Aim 3 demonstrated that paid advertisements and unpaid posts on Facebook are a useful way to reach women, 30 years of age or older, who have an increased risk for ovarian cancer. Using Facebook, the MAGENTA study team was able to increase the reach of study recruitment information, using targeted advertising, and collaborating with patient advocacy groups with an established Facebook following to both generate study notifications and boost signals from concurrent efforts, including news stories. Social media platform selection was an important factor in reaching the targeted audience, as was the use of high-quality imagery.

Images are important to social media users

Prior research has demonstrated that the use of certain structural components, such as hashtags, images and other media, improve the reach of social media messaging (Ordenes et al., 2018; Theiss et al., 2016). This is particularly true of the use of images or other visuals, which are associated with increased sharing over social media platforms (Ordenes et al., 2018; Theiss et al., 2016). This finding was consistent across each study aim. In aim 1, tweets with hashtags,

imagery, and “calls to action” were common. In aim 2, participants specifically remarked on the importance of images, noting that high quality, real images made them feel more confident in the source. Likewise, in aim 3, posts with high quality images trended better when compared to those without, particularly posts with images of patient advocates or scenery from targeted regions. These findings were consistent across Facebook and Twitter but may also be relevant to other social media platforms that allow users to utilize hashtags and to share media like images and video.

Users are able to assess quality of online health information

Prior studies have found that internet users are able to assess quality and accuracy of online health information, despite the plethora of inaccurate information readily available online (Cole et al., 2016). The results of aim 2 support these findings, noting that interview participants, all of whom were adult internet users who felt comfortable using the internet to find health related information, assessed unfamiliar online sources using several of the NIH’s recommendations for assessing online health information. Participants tended towards online sources that were familiar due to prior experience, personal referral, or geographic location, when available. Otherwise they sought out more information about the page and source behind it and looked at the Top Level Domain (TLD) of the website address, with a preference towards website types that were less likely to be published by a company trying to sell something, such as .org, .edu, and .gov (Chi et al., 2018). These findings were specific to online health information sources found over an internet search engine and may not be generalizable to online health sources encountered over social media, either by seeing a resource shared by a trusted friend or family member or through online information scanning. Future research should investigate if

social media users utilize similar assessment methods when they encounter a health resource over social media.

Leveraging Followers

Aim 1 results suggest that follower count is important when maximizing message reach. Accounts with a higher number of followers produced content that is shared, or retweeted, more often. This finding was relevant in Aim 2 participants as well, where users were more likely to see webpages with a higher number of recent visitors in their search results. In both examples, webpages and online content with a greater following (e.g. social media) or more visitors (e.g. other online webpages) arguably have more power to disseminate information effectively, or at the very least to push out information that will receive more impressions and have greater reach.

Aim 3 leveraged this observation in several ways. First, other Facebook pages and groups with a high follower count were tagged in unpaid posts, increasing the likelihood that individuals who followed or interacted with those existing Facebook pages and groups would see the unpaid post. Second, the MAGENTA study asked other Facebook pages and groups with a high follower count to post information about the MAGENTA study to their pages and groups directly. Finally, paid Facebook advertisements specifically targeted people who liked those Facebook pages or groups with high followers, using Facebook's audience targeting tool. Posts and online content incorporating these strategies tended to trend better online, with greater reach.

Another important aspect of this strategy, albeit more difficult to measure, is the social networking these interactions inspired. The MAGENTA Facebook page shared posts and resources from other Facebook pages and groups that had been asked to share MAGENTA study materials, in addition to publishing posts about the MAGENTA study. The goal behind this online social networking was to help increase the reach of other ovarian cancer related

messaging, on behalf of other organizations, in hopes that this may make it more likely that they would do the same, on behalf of MAGENTA, creating a mutually beneficial relationship. This was particularly true of Facebook pages and groups facilitated by ovarian cancer patient advocates, who published information about the MAGENTA study over their social media pages, effectively sharing this material across their established social network of followers. These relationships were key to recruitment success, illustrated by several recruitment spikes following efforts from patient advocates.

Lookalike audiences help improve targeted advertising

Social media platforms, like Facebook, are constantly changing and expanding their targeted advertising tools. This can pose a challenge to paid recruitment campaigns, both in implementation and assessment. For example, during the data collection period for aim 3, Facebook updated what impressions, defined as the number of times an ad was displayed on a screen, were included in campaign reporting. Prior to this update, Facebook would only report impressions that were budgeted for, but following this change they included impressions beyond that budget threshold. In other words, there are instances when Facebook has to increase the number of impressions an ad receives, in order to increase the number of results the advertisement receives. Changes like this may mitigate data collection accuracy around impressions. Changes to the advertisement tool may also help improve targeting and efficacy of a paid campaign.

Need for dynamic social media kits

Another important change to Facebook's advertising policies took place in response to the controversy around the role of social media advertisements surrounding the 2016 US Presidential election. Facebook now requires individuals hoping to post a politically themed

advertisement to undergo a more rigorous approval process before their ads will be posted. Advertisements that mention a specific person by name, even if they do not have a connection to politics, are flagged for further assessment. This change made approving advertisements for the MAGENTA study a more rigorous process, when posting ads containing participant stories or quotes from patient advocates connected to the study. It is likely that changes to online paid advertising over Facebook and other social media platforms will continue. This near constant flux of the OBA environment paves the way for innovative approaches to targeted recruitment but can also raise challenges around planning and tracking campaigns. As a result, media kits, like the one generated for the MAGENTA study, need to be flexible and forward thinking in order to adapt to the dynamic online environment.

Trouble with social bots

The social bot generated material collected over Twitter in aim 1 made it difficult to focus on human-generated content. This is not unique to Twitter, as social bots exist across all social media platforms. Material generated by social bots also cannot be dismissed, since human accounts share and interact with it, as observed in Aim 1 (Ferrara, 2018). Social bots were a concern as early as 2006, when researchers worried that they may be used to manipulate public opinion through social media (Ferrara, 2018; Howard, 2006). Social bots can manipulate social media content, emulating the activity of human users, but at a higher pace, and are getting better at not disclosing their true identity as they become more sophisticated (Varol et al., 2017B; Messias et al., 2013; Hwang et al., 2012). In other words, what social bots say matters to social media users, whether we want it to or not. More research is needed to better understand the impact social bot generated content has on public health information seeking and health communication efforts.

Future research

Social bots. Current research indicates that crowd sourcing or posting social media assessment queries to online forums where independent contractors or members of the public complete tasks for a fee, is an effective, albeit cost prohibitive, way to distinguish social bot accounts from human social media accounts. But this finding may not apply to online health information scanners, or individuals bumping in to health information over social media, where the more passive health information scanning may be more likely to occur. Future research should explore how social media users assess and interact with health information they see over their social media feeds, investigating if users are able to differentiate social bot generated content from human, and whether or not social bot generated content drives the type of content social media users interact with.

Cyberchondria. Participants in aim 2 demonstrated comfort in assessing information quality and accuracy, suggesting that the impact of poor-quality information is not a great concern. Across all interviews and scenarios, however, the issue of worry seemed to be more tied with symptoms. This observation aligns with prior research on cyberchondria, or online health-related information seeking fueled by anxiety about one's health, that amplifies an individual's health anxiety (te Peol et al., 2016). Future research should explore how online information seeking impacts the health anxiety of individuals with a confirmed genetic predisposition.

Platform choice. Aim 3 recruitment efforts resulted in a non-representative study population, despite the fact that Facebook boasts a relatively representative population of users (Chang et al., 2010). Prior work suggests that different groups have different response rates where research is concerned, meaning that targeted marketing, even over social media, is likely to leave certain groups underrepresented (Whitaker et al., 2017). It would be worthwhile to explore other social media platforms with a minority majority, for additional minority

recruitment opportunities. Future research should assess the efficacy of targeted minority recruitment across varying social media platforms, with current recommendations point to the importance of letting the target population inform the platform choice (Arigo et al., 2018).

There is ample evidence to indicate that social media is well positioned to address health communication deficits, including ovarian cancer genetic testing uptake. Social media enables researchers and health professionals to reach at risk individuals online, improving access to health information in the process. With the help of patient advocates, online communication tools like social media may also improves representation in research and health service provision, by widening messaging reach. Facebook in particular may be well suited to target the growing adult social media user population, however concerns with over representation of non-Hispanic white women and the over proliferation of social bot generated content across all social media platforms demands more research.

Appendices

Figure 1: Example Tweet, including an external link, and text content.

GYN Cancer | #GYNCSM liked

Steps Through @stepsthrough · Nov 29

"We needed more than chatting, sharing, laughter. We needed a language sufficient to speak about her **#ovariancancer** and what it meant. We needed support from someone who understood what was happening to us." Read: ow.ly/JPB530mNS4J #LifeWithOC #gynccsm @southpawclaire

Image or media

Social engagement

#hashtag

Call to action

@mentions

Figure 2: Study flow diagram detailing the flow of Aim 1, starting with data collection, to assessing data for eligibility criteria, to structural component and qualitative thematic assessment.

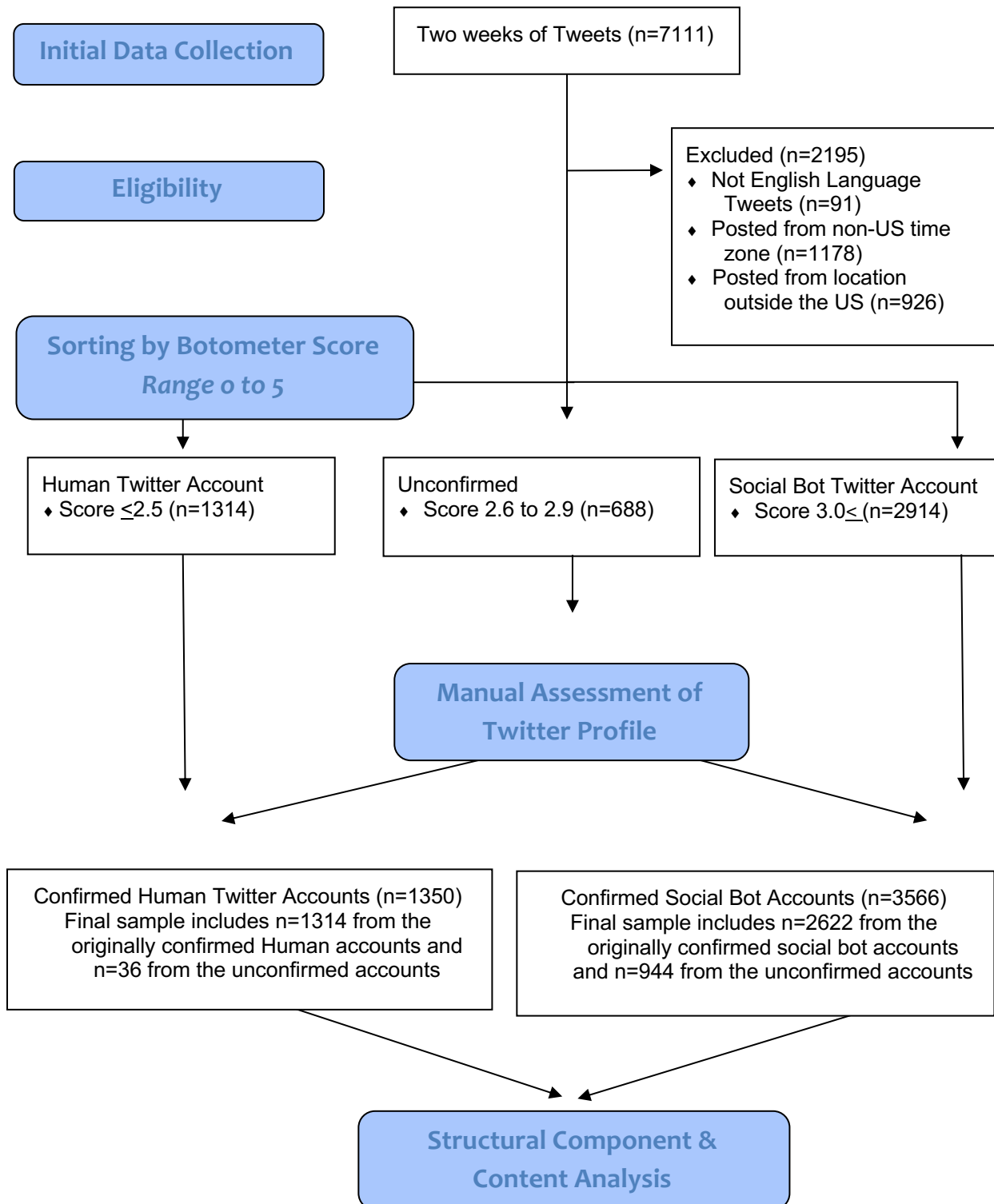


Figure 3. Sample of Botometer report.

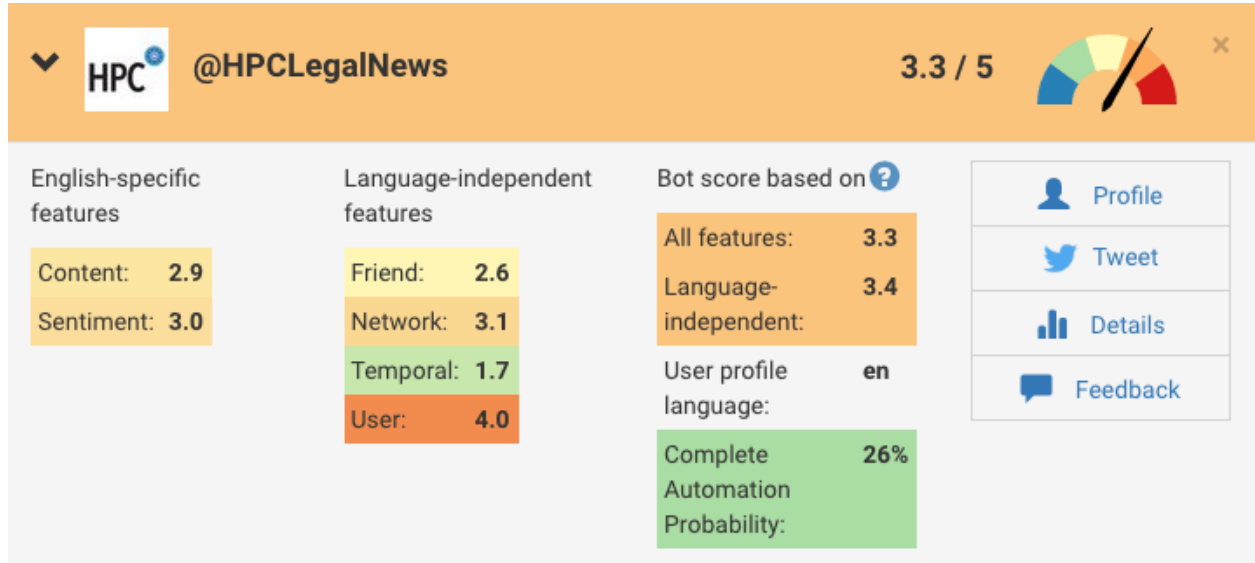


Figure 4. Plot of the eligible human and social bot generated Tweet count over the period from May 10 to May 23, 2017. The frequency of Tweeting, for both human-operated accounts and social bot-operated accounts peaked at 136 on May 18th for human-operated accounts, and 653 on May 19th, for social bot operated accounts. Individual bot accounts Tweeted 2.1 tweets per day on average, while human accounts produced an average of 1.1 tweets per day.

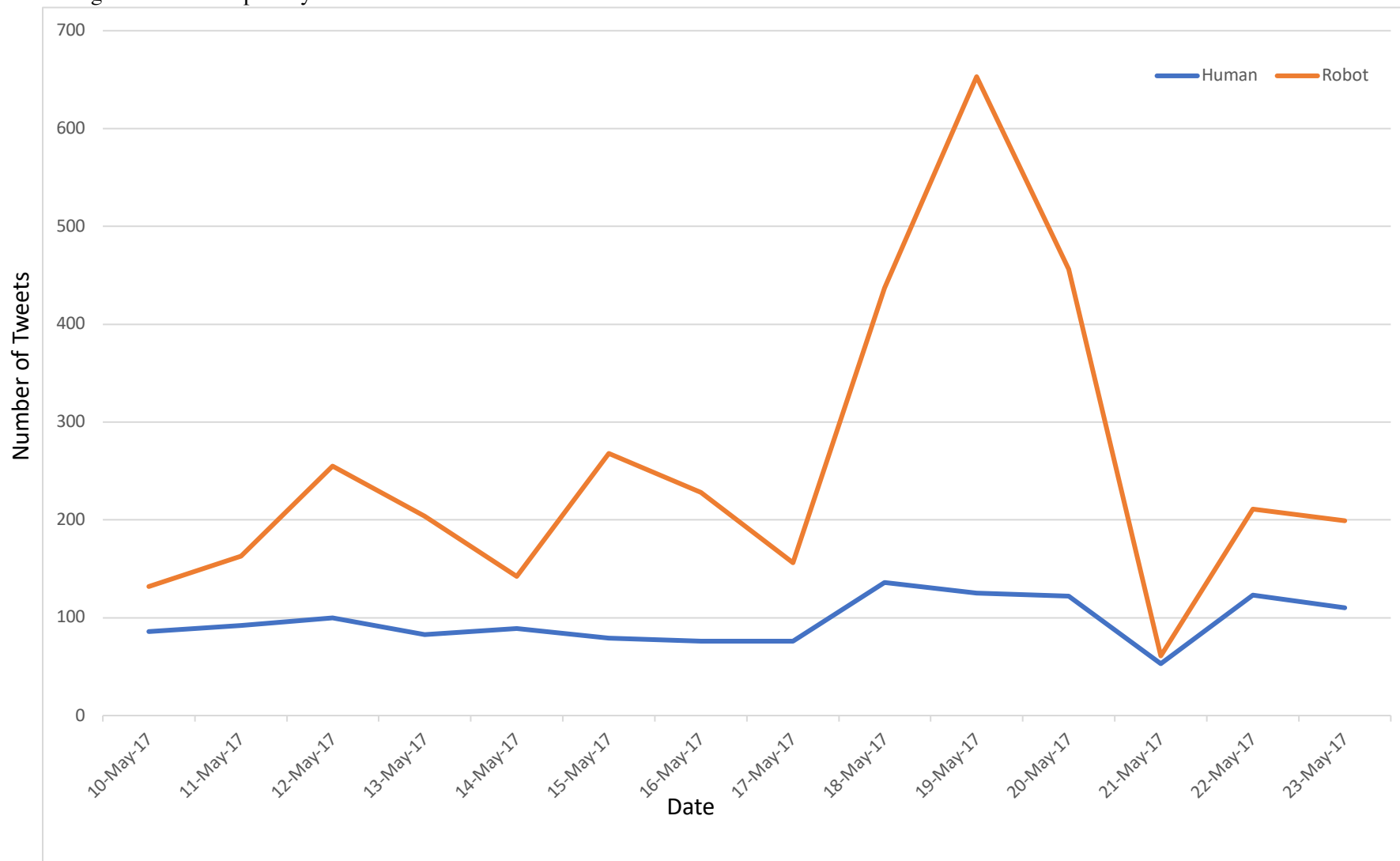


Table 1. Summary of total Tweets and user accounts, including number of followers and average Tweets per user, used across human-generated Tweets (human) and social bot-generated Tweets (social bot), observed over data collection period.

	Tweets		Users		Followers		Followers		Tweets/User	
	Human (n=1350)	Social Bot (n=3566)	Human (n=1226)	Social Bot (n=1621)	<i>Full data set</i>		<i>Top 10 Retweets</i>		Human	Social Bot
					Human	Social Bot	Human	Social Bot		
Mean	96.4	254.6	87.6	115.8	9131	6256	476310	6197	1.1	2.1
Median	90.5	207.5	83.0	116.5	614	2377	45658	1050	1.1	1.9
SD	23.7	157.9	20.5	38.2	67576	9536	959941	8933	0.0	0.8
Min	53	61	50	33	3	0	467	1050	1.1	1.2
Max	136	653	128	198	1433863	75281	3075280	24623	1.2	3.7
<i>p-value</i>	0.001*		0.022*		0.013*		0.139		<0.001*	

**Indicates a statistically significant value observed between human and social bot generated content.*

Table 2. Summary of the structural components observed across human-generated Tweets (human) and social bot-generated Tweets (social bot). Structural components include images (including media), hashtags, user mentions, URLs, and retweets.

	Images		Hashtags		User Mentions		URL		Retweets	
	Human	Bot	Human	Bot	Human	Bot	Human	Bot	Human	Bot
N (%)	800 (59)	1961 (55)	522 (39)	1328 (37)	935 (69)	2182 (61)	800 (59)	3287 (92)	686 (51)	1764 (49)
95% CI	0.8944		-1.0238		5.0229		30.2418		-1.1303	
	7.0679		5.0625		10.8996		35.7775		5.1243	
p-value	0.012*		0.196		<0.001*		<0.001*		0.211	
<i>*Indicates a statistically significant difference observed between human and social bot.</i>										

Table 3: Top ten most retweets from human operated twitter accounts.

Tweet	# of Retweets
RT @attorneyatlaw1: Ovarian Cancer Lawsuit: Thousands May Be Eligible for Compensation. No fees until you receive a settlement! https://t...	40
RT @OldSchool80s: May 20, 1989, Gilda Radner passed away at age 42 due to ovarian cancer. #RIP https://t.co/FhYpRGgAMx	26
RT @bobpockrass: Ovarian cancer survivor Calysta Bevier sings “Fight Song” - - which she sang on America’s Got Talent -- to open the Truex f...	22
RT @MIT: New technology could detect ovarian cancer months earlier than is now possible https://t.co/8ocCZdhm06 https://t.co/KtWwYwz3bc	20
RT @JoshMoney2: My beautiful mom needs help kicking cancer's butt! If you can't donate please share the link!! https://t.co/Cvur5fuEOh http...	19
RT @KattFunny: My incredible mother passed away last year from ovarian cancer. I hope I can love my child as much as she loved me. #Happy...	14
RT @BearsOutreach: Surprise mom w/ tix to the #BearsCareGala on 5/20 to meet her favorite #Bears & help fight breast & ovarian cancer: http...	13
RT @Orioles: Tonight we recognized Dr. Kala Visvanathan for her dedication to preventing both breast cancer & ovarian cancer. #BirdlandHero...	12
RT @WebMD: The strongest risk factor for ovarian cancer is age-- it's most likely to develop after menopause. More info: https://t.co/IxNK8...	10
RT @SoapDigest: GH/Sandy Rollman Ovarian Cancer Foundation Event https://t.co/Di8UfB8c1f @srocf @Robertpwatkins @WilliamdeVry1 @MauraWest @...	9

Table 4A: Summary of the thematic analysis of human-generated Tweets (human) and social bot-generated Tweets (social bot). Thematic analysis captured tone of Tweets, as well as common themes that appeared throughout the data, from risk factors related to ovarian cancer, to disease management and diagnosis, among other topics.

	Tone											
	<i>Awareness</i>		<i>Research</i>		<i>Support</i>		<i>Inspirational</i>		<i>Patient + Survivors</i>		<i>Fear-based language</i>	
	Human	Bot	Human	Bot	Human	Bot	Human	Bot	Human	Bot	Human	Bot
N	138	473	133	68	149	50	138	473	133	68	149	50
%	10.2	13.2	9.9	1.9	11.0	1.4	10.2	13.2	9.9	1.9	11.0	1.4
95% CI	-0.0936		6.3894		8.0164		6.2749		1.3880		32.6880	
	3.9538		9.7014		11.4550		10.2967		4.6029		36.4541	
P-value	0.060		<0.001		<0.001		<0.001		<0.001		<0.001	
<i>*Indicates a statistically significant value observed between human and social bot.</i>												

Table 4B: Summary of the thematic analysis of human-generated Tweets (human) and social bot-generated Tweets (social bot). Thematic analysis captured tone of Tweets, as well as common themes that appeared throughout the data, from risk factors related to ovarian cancer, to disease management and diagnosis, among other topics.

	Risk Factors						Disease Management & Diagnosis						Other	
	<i>Talcum Powder</i>		<i>Genetics</i>		<i>Other cancer</i>		<i>Diagnosis</i>		<i>Symptoms</i>		<i>Treatment</i>		<i>Sales</i>	
	Human	Bot	Human	Bot	Human	Bot	Human	Human	Bot	Human	Bot	Human	Bot	Human
N	55	51	59	27	190	244	57	55	51	59	27	190	244	57
%	4.1	1.4	4.4	0.8	14.1	6.8	4.2	4.1	1.4	4.4	0.8	14.1	6.8	4.2
95% CI	1.6087		2.5839		5.2704		1.9744		35.8932		-0.0620		6.5096	
	3.8787		4.8587		9.4477		4.2578		40.4119		4.0434		8.4634	
P-value	<0.001		<0.001		0.001		<0.001		<0.001		0.058		<0.001	

**Indicates a statistically significant value observed between human and social bot.*

<p>Table 5. Adapted from eHEALS questionnaire used in part one of the interview to assess eHealth literacy (Norman & Skinner, 2006), with additional questions driven by the CMIS model to assess prior experience with ovarian cancer, perceived risk and perceived severity and to get a sense of prior social media use. The majority of the questions were answered using a Likert scale, where 1=not useful at all/not important/strongly disagree, and 5=very useful/very important/strongly agree</p>
eHeals Q1. How useful do you feel the Internet is in helping you in making decisions about your health?
eHeals Q2. How useful do you feel online social media is in helping you in making decisions about your health? Online social media includes platforms like Facebook, Twitter, or another online support group.
eHeals Q3. How important is it for you to be able to access health resources on the Internet?
eHeals Q4. How important is it for you to be able to access health resources over social media, like Facebook, Twitter, or an online support group?
eHeals Q5. I know what health resources are available on the Internet
eHeals Q6. I know where to find helpful health resources on the Internet
eHeals Q7. I know how to find helpful health resources on the Internet
eHeals Q8. I know how to use the Internet to answer my questions about health
eHeals Q9. I know how to use the health information I find on the Internet to help me
eHeals Q10. I have the skills I need to evaluate the health resources I find on the Internet
eHeals Q11. I can tell high quality health resources from low quality health resources on the Internet
eHeals Q12. I feel confident in using information from the Internet to make health decisions
WQ1: Do you have social media accounts with any of the following social media platforms? <list examples: Facebook, Twitter, Instagram, etc.>
WQ2A: Do you have any family members who have been diagnosed with breast or ovarian cancer?
WQ2B. Do you have any friends who have been diagnosed with breast or ovarian cancer?
WQ3: I am at high risk of getting ovarian cancer in my life time.
WQ4: My genes put me at high risk of getting ovarian cancer in my life time.
WQ5: Ovarian cancer is a devastating disease.
WQ6: I can reduce my risk of ovarian cancer by engaging in preventative behaviors, like diet
WQ7: I know how to find information about ovarian cancer prevention.

Table 6: Interview task scenarios.

Scenario 1

Let's imagine that you just learned that breast and ovarian cancer run in your family. You want to learn more about what your personal risk of hereditary cancer is.

Using an internet browser of your choice, please show me how you might go about learning more what a family history of ovarian cancer means to your personal risk. While you are working, please talk through your thought process. Tell me about why you are doing what you are doing, what you notice, and what you are hoping to find.

Scenario 2

While browsing your Facebook feed, a story pops up, listing certain genetic mutations as a potential risk factor of ovarian cancer. The story explains that these same genetic mutations may also put you at higher risk of other types of cancer, including breast cancer. Show me how you would go about learning more about genetic mutations behind ovarian cancer.

Scenario 3

Now let's imagine that you received a positive test result for a *BRCA1* mutation. Knowing that certain types of mutations may be heritable, or "run in the family," you want to learn more about how this risk might impact other family members. While you are working, please talk through your thought process.

Scenario 4

Imagine you recently noticed a dull ache in your lower abdomen. Having recently saw a poster at your doctor's office that lists lower abdominal pain as a possible symptom for ovarian cancer, you decide to do a little more research to assess your concern, before contacting your doctor.

Using an internet browser of your choice, please show me how you might go about learning more about symptoms of ovarian cancer.

Table 7: Follow up interview questions.

FQ1: *Experience.* How do you feel about the idea of looking up information about _____ on the internet?

FQ2a: *Salience.* Do you think the internet is a useful resource for someone concerned about _____?

FQ2b: *Salience.* Would/has information you found online impacted your understanding/feelings towards/etc. ovarian cancer?

FQ2c: *Salience.* What role might online information play/has played in your ovarian cancer journey?

FQ3: *Beliefs.* Has information empowered you in any way, or made you feel like you had more control over your risk/diagnosis?

FQ4: *Beliefs.* Rate your agreement with the following statement: This information could be used to reduce my risk of ovarian cancer.

<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
Strongly disagree	Disagree	Neither agree, nor disagree	Agree	Strongly agree

FQ5: *Information-carrier characteristics.* Are there any specific markers you look for on a website or online source, that assure you it's a quality source?

FQ6: *Utility.* Does the information on this website meet your needs? Why/why not?

FQ7: *Utility.* Rate your agreement with the following statement: This source (specify online page or group) meets my informational needs.

<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5
Strongly disagree	Disagree	Neither agree, nor disagree	Agree	Strongly agree

Table 8: Mean scores for the adapted Individual eHEALS items and other section one questions, where 1=not useful/important at all/strongly disagree, and 5=very useful/important/strongly agree (n=12).

Statement	<u>Agreement score</u>
The internet is useful for helping me make decisions about my health	3.8
Online social media is useful for helping me make decisions about my health	1.8
It is important for me to be able to access health resources on the Internet.	4.7
It is important for me to be able to access health resources over social media.	2.1
I know what health resources are available on the Internet	3.8
I know where to find helpful health resources on the Internet	4.1
I know how to find helpful health resources on the Internet	4.2
I know how to use the Internet to answer my questions about health	4.4
I know how to use the health information I find on the Internet to help me	4.2
I have the skills I need to evaluate the health resources I find on the Internet	4.1
I can tell high quality health resources from low quality health resources on the Internet	4.2
I feel confident in using information from the Internet to make health decisions	3.6
I am at high risk of getting ovarian cancer in my life time.	3.5
My genes put me at high risk of getting ovarian cancer in my life time.	3.9
Ovarian cancer is a devastating disease.	4.6
I can reduce my risk of ovarian cancer by engaging in preventative behaviors, like diet and exercise.	4.4
I know how to find information about ovarian cancer prevention.	4.1

Figure 5: Consort flow diagram of study

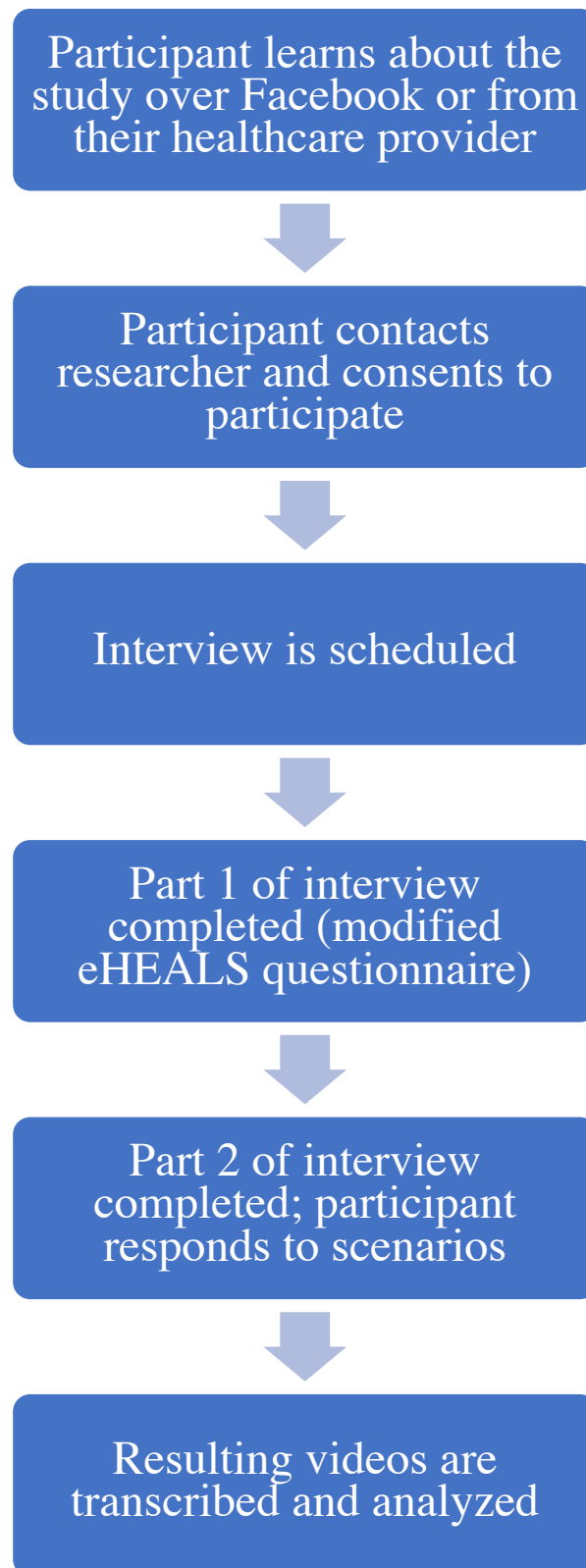



Figure 6: Example of an interactive online risk assessment quiz participants visited.

assessyourrisk™ by bright link

1 You 2 Your Health 3 Your Family 4 Results



Do you have Breasts & Ovaries?

YES **NO**

Why It Matters:
Every woman with breast and ovaries faces some risk of breast and ovarian cancer. On average, 1 in 8 women will develop breast cancer and 1 in 75 women will develop ovarian cancer.

This quiz was designed for women who still have both breasts and both ovaries. If you have had a breast and/or ovary removed, talk to your doctor to learn more about your personal cancer risk.

Sources: 1



Copyright © 2018 Bright Pink
Privacy Policy
| Terms and Conditions

Figure 7: Illustration of the online communication system utilized by the MAGENTA study.

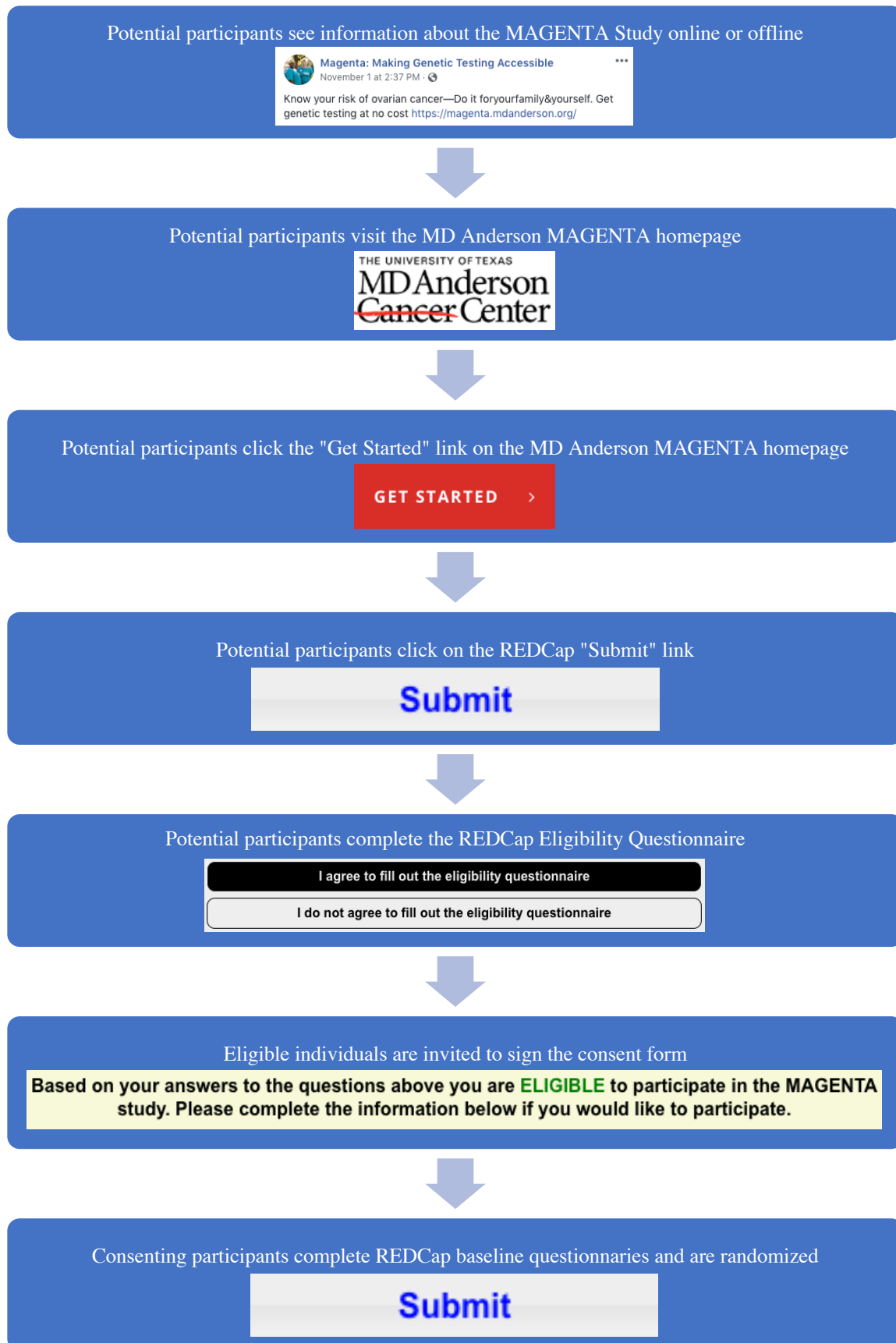
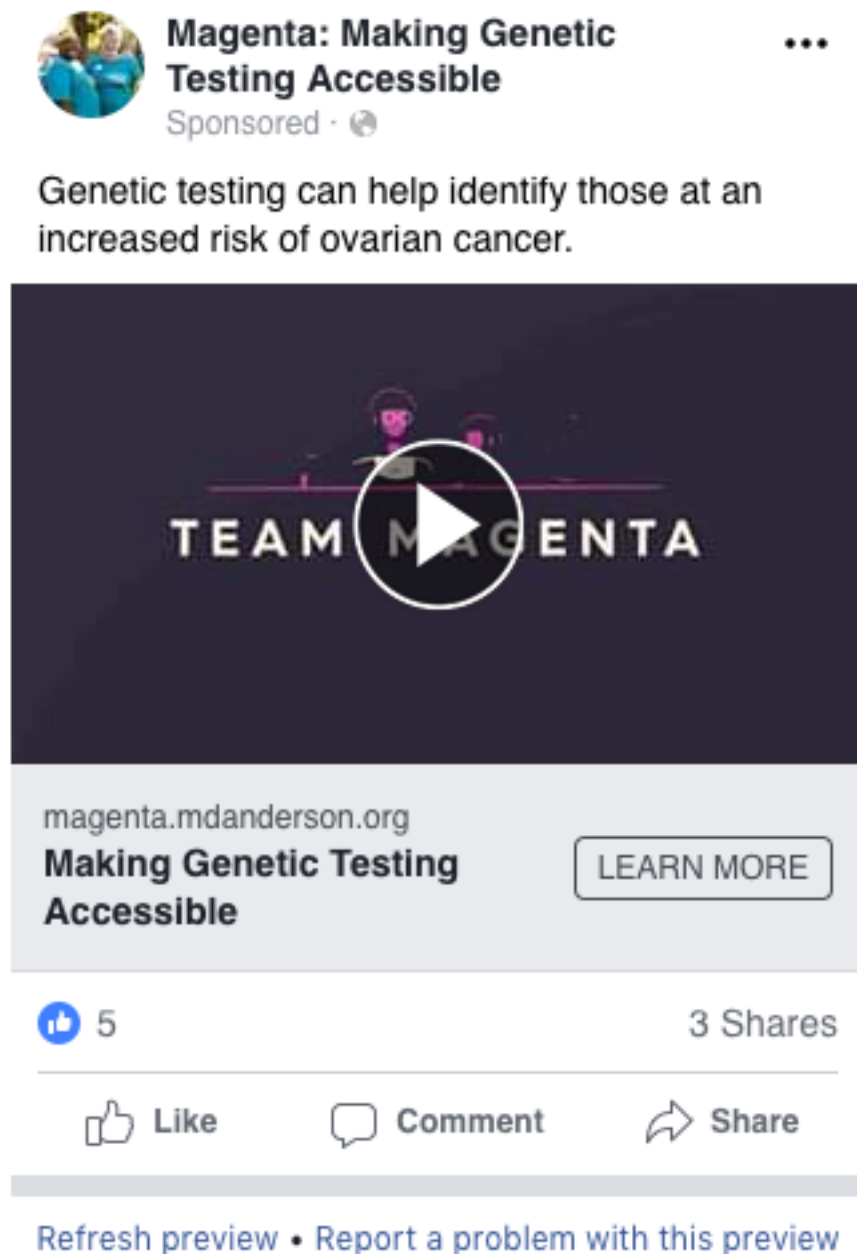



Figure 8: Example of paid Facebook advertisement used by the MAGENTA study. This paid advertisement included the study video.








Magenta: Making Genetic Testing Accessible Sponsored · 

Genetic testing can help identify those at an increased risk of ovarian cancer.

TEAM MAGENTA

magenta.mdanderson.org
Making Genetic Testing Accessible [LEARN MORE](#)

 5  3 Shares

 Like  Comment  Share

[Refresh preview](#) • [Report a problem with this preview](#)

Figure 9: Example of unpaid Facebook post published to the MAGENTA study Facebook page.



Magenta: Making Genetic Testing Accessible ...
Published by Tara Coffin [?]
Page Liked · November 1 · 🌐

Do you know someone battling ovarian cancer? Do you have a family member diagnosed with hereditary cancer? Are you worried about your risk? If you are a woman 30years of age or older, you may be eligible to learn your risk of ovarian cancer with genetic testing, at no cost. Testing is completed from the privacy of your home and will provide you with important information you may be able to use to protect your future. Take action today and find out if you're eligible for our study.<https://magenta.mdanderson.org/>

📍 Tag Photo 📍 Add Location ✎ Edit

👤 Jamie McLaughlin Crase 1 Share

👍 Like 💬 Comment ➦ Share 🌐

🌐 Write a comment...
Press Enter to post.

Table 9. Facebook analytic variables defined.

Term	Definition	Calculation Method
CPC (Cost per Link Click)	The average cost for a link click.	Total amount spent divided by link clicks.
CPM (Cost per 1000 impressions)	The average cost for 1,000 impressions (defined below).	Cost per 1000 impressions measures the total spent on an ad campaign, divided by the number of impressions, and multiplied by 1000.
CTR (Click through rate) (All)	The percentage of times people saw the ad and performed a click.	Total clicks divided by the number of impressions.
Link Click-Through Rate (CTR)	The percentage of times people saw the ad and clicked on the link.	Total link clicks, divided by the number of impressions.
Daily Budget	The average amount an individual indicated they were willing to spend on a specific ad per day.	N/A
Conversion Device	The device used by a person at the time they took an action based on the ad. Actions may include conversion events tracked by Facebook pixel, or certain on-Facebook actions. May include a desktop computer or a handheld mobile device, like an iPad, smart phone, etc.	N/A
Impressions	The number of times an ad was on screen.	N/A
Link Clicks	The number of clicks on links within the ad that led to an intended destination, on or off Facebook. For ads promoting the MD Anderson MAGENTA homepage, link clicks included click on the ad, that led to the homepage.	N/A
Lookalike Audiences	Lookalike audiences are created by Facebook to help advertisers reach people who “look like,” or are similar to an audience they would like to see the ad.	N/A
Objective	The objective for the campaign. The objective reflects advertisement goals.	N/A
Organic Leads	Organic leads include instances when someone sees an ad (or an unpaid post), and tags someone else who they think would be interested in it, and that person interacts with the ad as a result of that referral.	N/A
Page Engagement	Total number of actions that people took on the Facebook Page and its posts, as a result of an advertisement.	N/A
Page Likes	The number of likes of a Facebook Page received, as a result of an advertisement.	N/A

Placement	The location of where an ad is shown. Ads may show in a number of places, including: Facebook's mobile News Feed, desktop News Feed, right column, Instagram, Audience Network, Instant Articles and Messenger.	N/A
Post Comments	The total number of comments an ad receives.	The post comments metric counts all comments that people made on an ad while it was running.
Post Engagement	The total number of actions that people take involving an advertisement.	N/A
Post Likes	The number of likes on a post as a result of an ad.	N/A
Reach	The number of people who saw an ad at least once. This metric is different from impressions, which may include multiple views of an ad by the same people.	N/A
Results	The number of times an ad achieved an outcome, based on the assigned objective.	N/A
Unique clicks	Total number of individuals who performed a click.	N/A
Unique content views	Total number of individuals who viewed content.	N/A
Unique CTR (Click through rate)	The percentage of people who saw the ad and performed a unique click (all).	Unique clicks (all) divided by reach.
Unique link clicks	Total number of individuals who clicked on a link.	N/A
Video Plays	The total number of times a video starts to play. This includes all impressions when the video played, but does exclude video replays.	N/A

Source: Facebook, 2018B

Table 10. Description of Facebook paid campaign content and audience

Ad Description				Audience Description		
Ad Name	Ad Set ID	Objective	Media	Age	Geographic	Interests or Additional Targeting
Lookalike Audience Campaign	6099359114952	Conversions	Study video	30-65+	United States	Lookalike (US, 1%)-Website Traffic
CA Campaign 1	6096226885752	Conversions	Image	30-55	Salinas, CA	Breast cancer awareness, National Breast Cancer Foundation, National Ovarian Cancer Coalition
CA Campaign 2	6094137743552	Conversions	Study video	30 - 65+	Salinas, CA	Breast Cancer Care, Breast Cancer awareness, Ovarian Cancer National Alliance, National Ovarian Cancer Coalition or Ovarian Cancer Research Fund
CA Campaign 3	6094137490752	Conversions	Image	30 - 60	Downey, Pomona, Santa Ana, CA	Breast Cancer awareness, Telemundo or Univision
Cascade Testing 1	6099501882752	Conversions	Image	30-55	United States	Facing Our Risk of Cancer Empowered
Cascade Testing 2	6099501882752	Conversions	Image	30-55	United States	Facing Our Risk of Cancer Empowered
African American Target Campaign	6084365349552	Traffic	Study video	30 - 65+	Birmingham, Montgomery, AL; Miami Gardens, FL; Savannah, GA; New Orleans, LA; Baltimore, MD; Detroit, Flint, MI; Jackson, MS; Memphis, TN	N/A
NY Campaign	6083055715152	Traffic	Study video	30 - 65+	New York state	American Cancer Society, The Breast Cancer Research Foundation, National Breast Cancer Awareness Month, National Breast Cancer Foundation, Cancer Treatment Centers of America, Stand Up to Cancer, National Cancer Institute, Ovarian Cancer National Alliance, American Association for Cancer Research, Cancer research, Living Beyond Breast Cancer, Breast Cancer Campaign, Susan G. Komen 3-Day, National Ovarian Cancer Coalition, Ovarian Cancer Research Fund or Support The Fight Against Breast Cancer, Behaviors: Cancer Causes
BCOC Interest Campaign	6094525575952	Traffic	Image	30 - 65+	United States	
BCOC Interest Campaign	6079433429752	Engagement	Image	30 - 65+	Seattle, WA	The Breast Cancer Research Foundation, National Breast Cancer Awareness Month, Susan G. Komen for the Cure or Living Beyond Breast Cancer

BCOC Interest Campaign	6078968112152	Traffic	Image	30 - 65+	United States	N/A
BCOC Interest Campaign	6087376248352	Traffic	Study video	30 - 50	United States	American Cancer Society, Pink ribbon, The Breast Cancer Research Foundation, National Breast Cancer Awareness Month, Breast Cancer awareness, National Breast Cancer Foundation, Cancer Treatment Centers of America, Stand Up to Cancer, CANCER SUCKS!, National Cancer Institute, Susan G. Komen for the Cure, Ovarian Cancer National Alliance, Cancer research, Living Beyond Breast Cancer, Breast Cancer Campaign, Susan G. Komen 3-Day, National Ovarian Cancer Coalition, Ovarian Cancer Research Fund or Support The Fight Against Breast Cancer
BCOC Interest Campaign	6086437110552	Traffic	Study video	30 - 50	United States	American Cancer Society, Pink ribbon, The Breast Cancer Research Foundation, National Breast Cancer Awareness Month, Breast Cancer awareness, National Breast Cancer Foundation, Cancer Treatment Centers of America, Stand Up to Cancer, CANCER SUCKS!, National Cancer Institute, Susan G. Komen for the Cure, Ovarian Cancer National Alliance, Cancer research, Living Beyond Breast Cancer, Breast Cancer Campaign, Susan G. Komen 3-Day, National Ovarian Cancer Coalition, Ovarian Cancer Research Fund or Support The Fight Against Breast Cancer
BCOC Interest Campaign	6085654924552	Traffic	Study video	30 - 50	United States	American Cancer Society, Pink ribbon, The Breast Cancer Research Foundation, National Breast Cancer Awareness Month, Breast Cancer awareness, National Breast Cancer Foundation, Cancer Treatment Centers of America, Stand Up to Cancer, CANCER SUCKS!, National Cancer Institute, Susan G. Komen for the Cure, Ovarian Cancer National Alliance, Cancer research, Living Beyond Breast Cancer, Breast Cancer Campaign, Susan G. Komen 3-Day, National Ovarian Cancer Coalition, Ovarian Cancer Research Fund or Support The Fight Against Breast Cancer

Latino Target Campaign	6085309371752	Traffic	Image	30-65+	Brownsville, El Paso, Laredo, McAllen, TX; Hialeah, Florida; Downey, Oxnard, Pomona, Salinas, Santa Ana, CA	N/A
BCOC Interest Campaign	6081337489352	Traffic	Image	30-65+	United States	The Breast Cancer Research Foundation, Susan G. Komen for the Cure, Ovarian Cancer National Alliance, Living Beyond Breast Cancer, National Ovarian Cancer Coalition or Ovarian Cancer Research Fund
BCOC Interest Campaign	6080602441552	Traffic	Study video	30 - 65+	United States	The Breast Cancer Research Foundation, Susan G. Komen for the Cure, Ovarian Cancer National Alliance, Living Beyond Breast Cancer, National Ovarian Cancer Coalition or Ovarian Cancer Research Fund
WA Campaign	6091190202552	Conversions	Image	30-50	Seattle, WA	Essence (magazine); Latina (magazine)

Figure 10. Timeline of enrollment trends and recruitment events over time, and the number of responses received at different steps in recruitment activity.

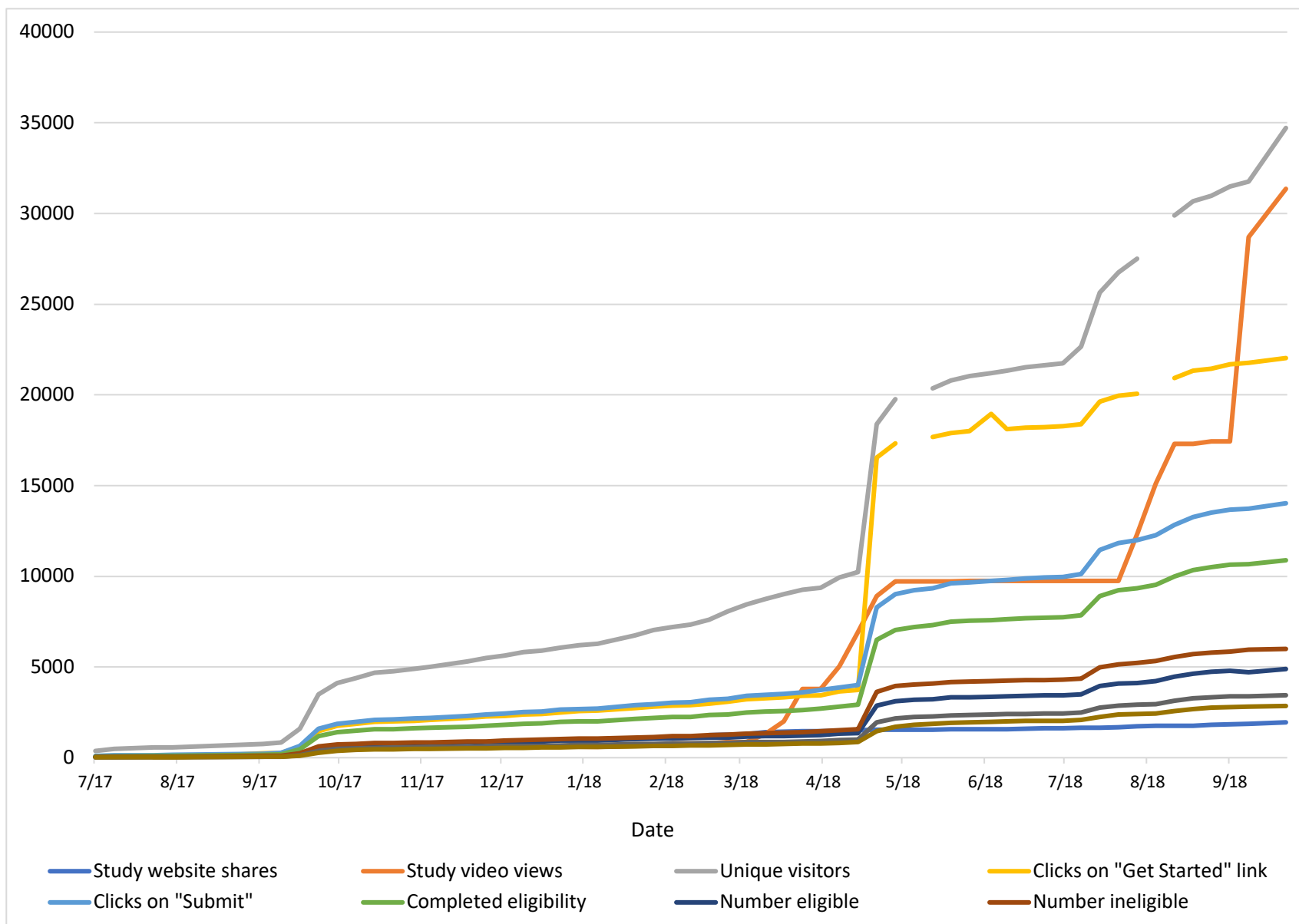


Table 11. Summary of results for all paid campaigns, along with engagement rates of different age demographics.

Ad Name	Objective	Age	Reach	Post Engagement	Page Engagement	Video Plays	Link Clicks	Results	Cost per Results
BCOC Interest Campaign 8	Traffic	ALL	2280	43	46	37	37	\$ 1.35	
		25-34	54	3	3	3	3	\$ 0.28	
		35-44	153						
		45-54	322	6	7	4	4	\$ 1.49	
		55-64	784	17	17	14	14	\$ 1.16	
		65+	967	17	19	16	16	\$ 1.56	
Latino Target Campaign 7	Traffic	ALL	11062	217	223	204	204	\$ 1.23	
		25-34	120	1	1				
		35-44	280	2	2	1	1	\$ 3.62	
		45-54	892	15	15	15	15	\$ 1.06	
		55-64	3547	62	65	57	57	\$ 1.20	
		65+	6223	137	140	131	131	\$ 1.23	
BCOC Interest Campaign 6	Traffic	ALL	7536	1397	1397	266	266	\$ 0.94	
		25-34	260	40	40	5	5	\$ 1.24	
		35-44	430	58	58	5	5	\$ 1.60	
		45-54	1160	159	159	37	37	\$ 1.02	
		55-64	2900	548	548	120	120	\$ 0.78	
		65+	2786	592	592	99	99	\$ 1.05	
BCOC Interest Campaign 5	Traffic	ALL	10436	2069	2069	280	280	\$ 1.07	
		25-34	146	15	15	1	1	\$ 3.50	
		35-44	414	30	30	3	3	\$ 2.66	
		45-54	1158	178	178	31	31	\$ 1.04	
		55-64	3698	645	645	105	105	\$ 0.94	
		65+	5020	1201	1201	140	140	\$ 1.12	
BCOC Interest Campaign 4	Traffic	ALL	34648	6536	6536	640	540	\$ 1.94	
		25-34	3496	427	427	53	37	\$ 2.18	
		35-44	13416	2050	2050	210	184	\$ 1.90	
		45-54	13672	2802	2802	283	240	\$ 2.03	
		55-64	1448	359	359	19	14	\$ 2.96	
		65+	2616	898	898	75	65	\$ 1.34	
BCOC Interest Campaign 9	Traffic	ALL	2050	351	351	20	20	\$ 2.00	
		25-34	33	4	4	1	1	\$ 0.37	
		35-44	88	13	13				
		45-54	258	27	27				
		55-64	821	141	141	12	12	\$ 1.21	
		65+	850	166	166	7	7	\$ 2.83	

African American Target Campaign	Traffic	ALL	9382	7308	7308	319	319	\$ 0.63	
		25-34	210	95	95	5	5	\$ 0.35	
		35-44	746	456	456	16	16	\$ 0.58	
		45-54	1312	873	873	39	39	\$ 0.57	
		55-64	2956	2276	2276	102	102	\$ 0.57	
		65+	4158	3608	3608	157	157	\$ 0.69	
Cascade Testing Campaign 1 and 2	Conversions	ALL	132480	3341	3342	3326	2514	\$ 0.17	
		25-34	20928	469	469	468	343	\$ 0.17	
		35-44	44992	1087	1087	1085	795	\$ 0.18	
		45-54	58944	1575	1576	1566	1211	\$ 0.16	
		55-64	7616	210	210	207	165	\$ 0.16	
WA Campaign	Conversions		20732	313	317	298	73	\$ 5.29	
		25-34	4592	61	61	61	15	\$ 4.87	
		35-44	10588	148	150	142	34	\$ 5.56	
		45-54	5552	104	106	95	24	\$ 5.17	
CA Campaign 3	Conversions		35976	2022	2022	2022	1692	\$ 0.21	
		25-34	6176	323	323	323	283	\$ 0.19	
		35-44	12320	687	687	687	570	\$ 0.21	
		45-54	10944	602	602	602	484	\$ 0.22	
		55-64	6536	410	410	410	355	\$ 0.21	
NY Campaign	Traffic	ALL	395	186	186	14	14	\$ 0.53	
		25-34	9	3	3				
		35-44	30	14	14	1	1	\$ 0.39	
		45-54	65	17	17	2	2	\$ 0.55	
		55-64	125	63	63	8	8	\$ 0.30	
		65+	166	89	89	3	3	\$ 1.16	
Lookalike Audience Campaign	Conversions	ALL	19240	7142	7142	12244	1130	880	\$ 0.40
		25-34	1424	336	336	742	70	57	\$ 0.41
		35-44	3844	823	823	1889	178	145	\$ 0.40
		45-54	4376	1280	1280	2526	263	196	\$ 0.40
		55-64	5812	2070	2070	3529	301	227	\$ 0.45
		65+	3784	2633	2633	3558	318	255	\$ 0.35
BCOC Interest Campaign 1	Traffic		1587						
		25-34	140	3	3		1	1	\$ 2.33
		35-44	420	3	3		1	1	\$ 6.36
		45-54	497	10	10		3	3	\$ 3.10
		55-64	380	14	14		2	2	\$ 4.20

		65+	150	1	2				
BCOC Interest Campaign 2	Engagement		334	55	55		55	\$ 0.27	
		25-34	13	2	2		2	\$ 0.15	
		35-44	35	7	7		7	\$ 0.22	
		45-54	64	4	4		4	\$ 0.55	
		55-64	93	18	18		18	\$ 0.21	
		65+	129	24	24		24	\$ 0.30	
BCOC Interest Campaign 3	Traffic		746	10	10		10	\$ 1.50	
		25-34	22						
		35-44	78	1	1		1	\$ 0.94	
		45-54	111	1	1		1	\$ 2.02	
		55-64	254	2	2		2	\$ 2.21	
		65+	281	6	6		6	\$ 1.23	
CA Campaign 2	Conversions		35752	21444	21444	7548	3147	2530	\$ 0.29
		25-34	4480	2055	2055	647	344	287	\$ 0.26
		35-44	8472	3955	3955	1311	681	512	\$ 0.29
		45-54	7488	4359	4359	1521	663	546	\$ 0.28
		55-64	8032	5323	5323	1928	771	629	\$ 0.29
		65+	7280	5752	5752	2141	688	556	\$ 0.32
CA Campaign 1	Conversions		23872	1637	1637		1637	1406	\$ 0.35
		25-34	5536	335	335		335	342	\$ 0.29
		35-44	9344	610	610		610	486	\$ 0.36
		45-54	8064	624	624		624	528	\$ 0.38
		55-64	928	68	68		68	50	\$ 0.46

Table 12. Summary of enrollment and randomization data.

Step in enrollment protocol	Total	% of total Unique Visitors
MAGENTA Facebook page Content		
Reach	407,769	N/A
Engagement (Clicks, reactions, comments, Facebook content shares)	57,248	N/A
Study website shares	1,948	N/A
Study video views	31,358	N/A
MAGENTA MD Anderson Webpage		
Unique visitors	34,715	-
Clicks on “Get Started” Link	22,029	63.5%
REDCap Introductory Message		
Clicks on “Submit” Link	14,025	40.4%
REDCap Eligibility		
Completed eligibility questionnaire	10,883	31.3%
Number eligible	4,887	14.1%
Number ineligible	5,996	17.3%
Study Informed Consent		
Signed consent	3,442	9.9%
REDCap Questionnaires		
Number of participants who complete the questionnaires (REDCap)	2,851	8.2%
Number ppts randomized To one of four conditions	2,851	8.2%

References

- Abrahamson, J. A., & Fisher, K. E. (2007). “What”’s past is prologue’: towards a general model of lay information mediary behaviour. *Information Research*, 12, 1–21.
- Abramson, K., Keefe, B., & Chou, W. Y. S. (2015). Communicating about cancer through Facebook: a qualitative analysis of a breast cancer awareness page. *Journal of Health Communication*, 20(2), 237-243.
- Acheson, L. S., Stange, K. C., & Zyzanski, S. (2005). Clinical genetics issues encountered by family physicians. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, 7(7), 501–508. <https://doi.org/10.1097/01.gim.0000177418.24176.9b>
- Adams, S. A. (2010)A. Blog-based applications and health information: Two case studies that illustrate important questions for Consumer Health Informatics (CHI) research. *International Journal of Medical Informatics*, 79(6). <https://doi.org/10.1016/j.ijmedinf.2008.06.009>
- Adams, S. A. (2010)B. Revisiting the online health information reliability debate in the wake of “ web 2.0”: An inter-disciplinary literature and website review. *International Journal of Medical Informatics*. <https://doi.org/10.1016/j.ijmedinf.2010.01.006>
- Aghaei, S., Nematbakhsh, M. A., & Farsani, H. K. (2012). Evolution of the world wide web: From WEB 1.0 TO WEB 4.0. *International Journal of Web & Semantic Technology*, 3(1), 1.
- Allem, J. P., Ferrara, E., Uppu, S. P., Cruz, T. B., & Unger, J. B. (2017). E-cigarette surveillance with social media data: social bots, emerging topics, and trends. *JMIR public health and surveillance*, 3(4).
- Allen, K. (2013). The average Facebook user is getting older—and more masculine. *Regan’s PR Daily*.
- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin*. 2018;68:7-30.
- Ancker, J. S., Carpenter, K. M., Greene, P., Hoffman, R., Kukafka, R., Marlow, L. A., ... & Quillin, J. M. (2009). Peer-to-peer communication, cancer prevention, and the internet. *Journal of health communication*, 14(S1), 38-46.
- Anderson, B., McLosky, J., Wasilevich, E., Lyon-Callo, S., Duquette, D., & Copeland, G. (2012). Barriers and facilitators for utilization of genetic counseling and risk assessment services in young female breast cancer survivors. *Journal of Cancer Epidemiology*. <https://doi.org/10.1155/2012/298745>
- Arigo, D., Pagoto, S., Carter-Harris, L., Lillie, S. E., & Nebeker, C. (2018). Using social media for health research: Methodological and ethical considerations for recruitment and intervention delivery. *DIGITAL HEALTH*, 4, 2055207618771757.
- Arora, N. K., Hesse, B. W., Rimer, B. K., Viswanath, K., Clayman, M. L., & Croyle, R. T. (2008). Frustrated and confused: The American public rates its cancer-related information-seeking experiences. *Journal of General Internal Medicine*, 23(3), 223–228. <https://doi.org/10.1007/s11606-007-0406-y>
- Aspinwall, L. G., Taber, J. M., Kohlmann, W., Leaf, S. L., & Leachman, S. A. (2014). Perceived risk following melanoma genetic testing: A 2-year prospective study distinguishing subjective estimates from recall. *Journal of Genetic Counseling*, 23(3), 421–437. <https://doi.org/10.1007/s10897-013-9676-1>

- Bajardi, P., Paolotti, D., Vespignani, A., Eames, K., Funk, S., Edmunds, W. J., ... Richiardi, L. (2014). Association between recruitment methods and attrition in internet-based studies. *PLoS ONE*, 9(12). <https://doi.org/10.1371/journal.pone.0114925>
- Baltar, F., & Brunet, I. (2012). Social research 2.0: virtual snowball sampling method using Facebook. *Internet research*, 22(1), 57-74.
- Bánki, F., Thomas, S. J., Main, B. G., & Waylen, A. E. (2016). Communication of information about oral and oropharyngeal cancer: the quality of online resources. *Oral Surgery*, n/a-n/a. <https://doi.org/10.1111/ors.12208>
- Basch, E. M., Thaler, H. T., Shi, W., Yakren, S., & Schrag, D. (2004). Use of information resources by patients with cancer and their companions. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 100(11), 2476-2483.
- Batterham, P. J. (2014). Recruitment of mental health survey participants using Internet advertising: Content, characteristics and cost effectiveness. *International Journal of Methods in Psychiatric Research*, 23(2), 184–191. <https://doi.org/10.1002/mpr.1421>
- Bauermeister, J. a, Zimmerman, M. a, Johns, M. M., Glowacki, P., Stoddard, S., & Volz, E. (2012). Innovative recruitment using online networks: lessons learned from an online study of alcohol and other drug use utilizing a web-based, respondent-driven sampling (webRDS) strategy. *Journal of Studies on Alcohol and Drugs*, 73(5), 834–8. <https://doi.org/10.15288/jsad.2012.73.834>
- Bean, S. J. (2011). Emerging and continuing trends in vaccine opposition website content. *Vaccine*, 29(10), 1874–1880. <https://doi.org/10.1016/j.vaccine.2011.01.003>
- Bessi, A., & Ferrara, E. (2016). Social bots distort the 2016 US Presidential election online discussion.
- Beyer, H., & Holtzblatt, K. (1997). Contextual Design: A Customer-Centered Approach to Systems Designs. *Contextual Design A CustomerCentered Approach to Systems Designs*. Retrieved from <http://www.amazon.com/dp/1558604111>
- Brickman Bhutta, C. (2012). Not by the Book: Facebook as a Sampling Frame. *Sociological Methods & Research*, 41(1), 57–88. <https://doi.org/10.1177/0049124112440795>
- Blomberg, J., & Giacomi, J. (1993). Participatory design: In D. Schuler & A. Namioka (Eds.), *Participatory Design: Principles and Practices*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Bonevski, B., Randell, M., Paul, C., Chapman, K., Twyman, L., Bryant, J., ... Hughes, C. (2014). Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. *BMC Medical Research Methodology*, 14(1), 42. <https://doi.org/10.1186/1471-2288-14-42>
- Bosker, B. (2011). The Average Facebook User: Stats show how we use the site. *Huffington Post*. Retrieved September 27, 2018 from https://www.huffingtonpost.com/2011/11/19/the-average-facebook-user_n_1102902.html
- Bowen, A., Williams, M., & Horvath, K. (2004). Using the internet to recruit rural MSM for HIV risk assessment: Sampling issues. *AIDS and Behavior*, 8(3), 311–319. <https://doi.org/10.1023/B:AIBE.0000044078.43476.1f>
- Bradbury, A. R., Patrick-Miller, L., Egleston, B., Sands, C. B., Li, T., Schmidheiser, H., ... Daugherty, C. K. (2010). Parent opinions regarding the genetic testing of minors for BRCA1/2. *Journal of Clinical Oncology*, 28(21), 3498–3505. <https://doi.org/10.1200/JCO.2009.27.2971>

- Brandt, R., Ali, Z., Sabel, A., McHugh, T., & Gilman, P. (2008). Cancer genetics evaluation: barriers to and improvements for referral. *Genetic Testing*, 12(1), 9–12.
<https://doi.org/10.1089/gte.2007.0036>
- Briel, M., Olu, K. K., von Elm, E., Kasenda, B., Alturki, R., Agarwal, A., ... & Schandelmaier, S. (2016). A systematic review of discontinued trials suggested that most reasons for recruitment failure were preventable. *Journal of clinical epidemiology*, 80, 8-15.
- Bruns, A., & Stieglitz, S. (2013). Towards more systematic Twitter analysis: Metrics for tweeting activities. *International Journal of Social Research Methodology*, 16, 91–108.
- Bruns, A., & Burgess, J. (2012). Researching news discussion on Twitter. *Journalism Studies*, 13, 801–814.
- Bull, S. S., Breslin, L. T., Wright, E. E., Black, S. R., Levine, D., & Santelli, J. S. (2011). Case study: An ethics case study of HIV prevention research on facebook: The just/us study. *Journal of Pediatric Psychology*. <https://doi.org/10.1093/jpepsy/jsq126>
- Bullock, B. N., Hotho, A., & Stumme, G. (2018). Accessing Information with Tags: Search and Ranking. In *Social Information Access* (pp. 310-343). Springer, Cham.
- Bunnik, E. M., Janssens, A. C. J. W., & Schermer, M. H. N. (2015). Personal utility in genomic testing: is there such a thing? *Journal of Medical Ethics*, 41(4), 322–326.
<https://doi.org/10.1136/medethics-2013-101887>
- Burton, S. H., Tanner, K. W., Giraud-Carrier, C. G., West, J. H., & Barnes, M. D. (2012). Right time, right place" health communication on twitter: Value and accuracy of location information. *Journal of Medical Internet Research*, 14(6).
<https://doi.org/10.2196/jmir.2121>
- Campbell, I. (2007). Chi-squared and Fisher–Irwin tests of two-by-two tables with small sample recommendations. *Statistics in medicine*, 26(19), 3661-3675.
- Campeau, P. M., Foulkes, W. D., & Tischkowitz, M. D. (2008). Hereditary breast cancer: new genetic developments, new therapeutic avenues. *Human genetics*, 124(1), 31-42.
- Caplan, A., & Friesen, P. (2017). Health disparities and clinical trial recruitment: Is there a duty to tweet?. *PLoS biology*, 15(3), e2002040.
- Carlsson, M. E. (2009). Cancer patients seeking information from sources outside the health care system: Change over a decade. *European Journal of Oncology Nursing*, 13(4), 304–305.
<https://doi.org/10.1016/j.ejon.2009.03.005>
- Carpenter, M. J., Sterba, K. R., Boatright, A. S., & West, R. (2011). “Closet” quit attempts: Prevalence, correlates and association with outcome. *Addiction*, 106(12), 2214–2220.
<https://doi.org/10.1111/j.1360-0443.2011.03538.x>
- Carter-Harris, L., Bartlett, Ellis R., Warrick, A., & Rawl, S. (2016) Beyond traditional newspaper advertisement: leveraging Facebook-Targeted Advertisement to recruit long-term smokers for research. *Journal of Med Internet Res*. 18(6).
- Case, D. O. (2012). *Looking for information: A survey of research on information seeking, needs and behavior* (3rd ed.). Bingley, UK: Emerald Group Publishing.
- Chang, J., Rosenn, I., Backstrom, L., & Marlow, C. (2010). ePluribus: Ethnicity on Social Networks. *ICWSM*, 10, 18-25.
- Charmaz, K. (2006). *Constructing grounded theory: a practical guide through qualitative analysis*. Book (Vol. 10). <https://doi.org/10.1016/j.lisr.2007.11.003>
- Charters, E. (2003). The use of think-aloud methods in qualitative research an introduction to think-aloud methods. *Brock Education Journal*, 12(2), 68–82.
<https://doi.org/10.1080/02602938.2010.496532>

- Chen, X., & Siu, L. (2001). Impact of the media and the Internet on oncology: Survey of cancer patients and oncologists in Canada. *Journal of Clinical Oncology*, 19(23), 4291–4297.
- Chew, C. M., & Eysenbach, G. (2010). Pandemics in the Age of Twitter : A Content Analysis of the 2009 H1N1 Outbreak by A Content Analysis of the 2009 H1N1 Outbreak. *PloS one*. Retrieved from https://tspace.library.utoronto.ca/bitstream/1807/25454/1/Chew_Cynthia_M_201011_MS_c_thesis.pdf
- Chi, Y., He, D., Han, S., & Jiang, J. (2018, March). What Sources to Rely on:: Laypeople's Source Selection in Online Health Information Seeking. In *Proceedings of the 2018 Conference on Human Information Interaction&Retrieval* (pp. 233-236). ACM.
- Chiang, A. L., Vartabedian, B., & Spiegel, B. (2016). Harnessing the Hashtag: A Standard Approach to GI Dialogue on Social Media. *The American Journal of Gastroenterology*.
- Chou, W. Y. S., Liu, B., Post, S., & Hesse, B. (2011). Health-related Internet use among cancer survivors: data from the Health Information National Trends Survey, 2003–2008. *Journal of Cancer Survivorship*, 5(3), 263-270.
- Cole, J., Watkins, C., & Kleine, D. (2016). Health advice from internet discussion forums: how bad is dangerous?. *Journal of medical Internet research*, 18(1).
- Comm, J. (2009). Twitter Power: How to Dominate Your Market One Tweet at a Time. *Journal of Chromatography A* (Vol. 776). [https://doi.org/10.1016/S0021-9673\(97\)00606-7](https://doi.org/10.1016/S0021-9673(97)00606-7)
- Cooper, L. Z., Larson, H. J., & Katz, S. L. (2008). Protecting public trust in immunization. *Pediatrics*, 122(1), 149–153. <https://doi.org/DOI 10.1542/peds.2008-0987>
- Correa, T., Willard, H. A., & de Zúñiga, H. (2010). Who interacts on the Web?: The intersection of users' personality and social media use. *Computers in Human Behavior*, 26(2), 247–253.
- Crocco, a. G., Villasis-Keever, M., & Jadad, a. R. (2002). Two Wrongs Don't Make a Right: Harm Aggravated by Inaccurate Information on the Internet. *Pediatrics*, 109(3), 522–523. <https://doi.org/10.1542/peds.109.3.522>
- Curtis, B. (2012). Understanding Tailored Internet Smoking Cessation Messages: A Reasoned Action Approach. *The ANNALS of the American Academy of Political and Social Science*, 640(1), 136–149. <https://doi.org/10.1177/0002716211423642>
- Curtis, B. L. (2014). Social networking and online recruiting for HIV research: Ethical challenges. *Journal of Empirical Research on Human Research Ethics: An International Journal*, 9(1), 58–70. <https://doi.org/10.1525/jer.2014.9.1.58>
- Czaja, R., Manfredi, C., & Price, J. (2003). The determinants and consequences of information seeking among cancer patients. *Journal of Health Communication*, 8(6), 529–562. <https://doi.org/10.1080/10810730390250470>
- Davis, C. A., Varol, O., Ferrara, E., Flammini, A., & Menczer, F. (2016, April). Botornot: A system to evaluate social bots. In *Proceedings of the 25th International Conference Companion on World Wide Web* (pp. 273-274). International World Wide Web Conferences Steering Committee.
- Derdiarian, A. K. (1986). Informational needs of recently diagnosed cancer patients. *Nurs Res*, 35(5), 276–281.
- Dew, A., Khan, S., Babinski, C., Michel, N., Heffernan, M., Stephan, S., ... Bergan, R. (2013). Recruitment strategy cost and impact on minority accrual to a breast cancer prevention trial. *Clinical Trials* (London, England), 10(2), 292–9. <https://doi.org/10.1177/1740774512471452>

- Dizon, D. S., Graham, D., Thompson, M. A., Johnson, L. J., Johnston, C., Fisch, M. J., & Miller, R. (2012). Practical guidance: the use of social media in oncology practice. *Journal of oncology practice*, 8(5), e114-e124.
- Doherty, P., Stott, A., & Kinder, K. (2004). Delivering services to hard to reach families in on track areas: definition, consultation and needs assessment, 14. Retrieved from <http://webarchive.nationalarchives.gov.uk/20110220105210/rds.homeoffice.gov.uk/rds/pdfs2/dpr15.pdf>
- Duggan, M., Ellison, N. B., Lampe, C., Lenhart, A., & Madden, M. (2015). Social media update 2014. Pew Research Center, 9.
- Duggan, M. (2015). The Demographics of Social Media Users. Retrieved from <http://www.pewinternet.org/2015/08/19/the-demographics-of-social-media-users/>
- Dutta-Bergman, M. J. (2004). Primary sources of health information: comparisons in the domain of health attitudes, health cognitions, and health behaviors. *Health Commun*, 16(3), 273–288. https://doi.org/10.1207/S15327027HC1603_1
- Elyashar, A., Fire, M., Kagan, D., & Elovici, Y. (2013, August). Homing socialbots: intrusion on a specific organization's employee using Socialbots. In *Advances in Social Networks Analysis and Mining (ASONAM)*, 2013 IEEE/ACM International Conference on (pp. 1358-1365). IEEE.
- Eysenbach, G., & Köhler, C. (2002). How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. *Bmj*, 324(7337), 573-577.
- Eysenbach, G., Powell, J., Kuss, O., & Sa, E.-R. (2002). Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. *JAMA : The Journal of the American Medical Association*, 287(20), 2691–700. <https://doi.org/10.1001/jama.287.20.2691>
- Eysenbach, G. (2007). From intermediation to disintermediation and apomediation: new models for consumers to access and assess the credibility of health information in the age of Web2. 0. In *Medinfo 2007: Proceedings of the 12th World Congress on Health (Medical) Informatics; Building Sustainable Health Systems* (p. 162). IOS Press.
- Eysenbach, G. (2003). The impact of the Internet on cancer outcomes. *CA: A Cancer Journal for Clinicians*, 53(6), 356–71. <https://doi.org/10.3322/canjclin.53.6.356>
- Facebook. (2018A). About Facebook Pixel. Retrieved September 3, 2018, from <https://www.facebook.com/business/help/742478679120153>.
- Facebook. (2018B). Glossary of Ad Terms. Retrieved September 5, 2018, from <https://www.facebook.com/business/help/447834205249495>
- Fallowfield, L. J. (1990). Psychosocial adjustment after treatment for early breast cancer. *Oncology (Williston Park, N.Y.)*, 4(4), 89-97-98, 100.
- Farmer, a D., Bruckner Holt, C. E. M., Cook, M. J., & Hearing, S. D. (2009). Social networking sites: a novel portal for communication. *Postgraduate Medical Journal*, 85(1007), 455–9. <https://doi.org/10.1136/pgmj.2008.074674>
- Feather, J. S., Howson, M., Ritchie, L., Carter, P. D., Parry, D. T., & Koziol-McLain, J. (2016). Evaluation methods for assessing users' psychological experiences of web-based psychosocial interventions: A systematic review. *Journal of Medical Internet Research*, 18(6). <https://doi.org/10.2196/jmir.5455>
- Febbraro, T., Robison, K., Wilbur, J. S., Laprise, J., Bregar, A., Lopes, V., ... Stuckey, A. (2015). Adherence patterns to National Comprehensive Cancer Network (NCCN)

- guidelines for referral to cancer genetic professionals. *Gynecologic Oncology*, 138(1), 109–14. <https://doi.org/10.1016/j.ygyno.2015.04.029>
- Fenner, Y., Garland, S. M., Moore, E. E., Jayasinghe, Y., Fletcher, a, Tabrizi, S. N., ... Wark, J. D. (2012). Web-based recruiting for health research using a social networking site: an exploratory study. *Journal of Medical Internet Research*, 14(1), e20. <https://doi.org/10.2196/jmir.1978>; 10.2196/jmir.1978
- Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, 59(7), 96-104.
- Ferrara, E. (2015). Manipulation and abuse on social media by emilio ferrara with ching-man au yeung as coordinator. *ACM SIGWEB Newsletter*, (Spring), 4.
- Ferrara, E. (2015). Manipulation and abuse on social media by emilio ferrara with ching-man au yeung as coordinator. *ACM SIGWEB Newsletter*, (Spring), 4.
- Ferrara, E. (2018). Measuring social spam and the effect of bots on information diffusion in social media. In *Complex Spreading Phenomena in Social Systems* (pp. 229-255). Springer, Cham.
- Finch, A., Metcalfe, K. A., Chiang, J. K., Elit, L., McLaughlin, J., Springate, C., ... Narod, S. A. (2011). The impact of prophylactic salpingo-oophorectomy on menopausal symptoms and sexual function in women who carry a BRCA mutation. *Gynecologic Oncology*, 121(1), 163–168. <https://doi.org/10.1016/j.ygyno.2010.12.326>
- Finch, A. P. M., Lubinski, J., Møller, P., Singer, C. F., Karlan, B., Senter, L., ... Narod, S. a. (2014). Impact of Oophorectomy on Cancer Incidence and Mortality in Women With a BRCA1 or BRCA2 Mutation. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, 1–8. <https://doi.org/10.1200/JCO.2013.53.2820>
- Foltz, A., & Sullivan, J. (1996). Reading level, learning presentation preference, and desire for information among cancer patients. *Journal of Cancer Education*, 11(1), 32–38. <https://doi.org/10.1080/08858199609528389>
- Fox, S., & Duggan, M. (2013). *Health online 2013*. Washington, DC: Pew Internet & American Life Project, 1.
- Fox, S., & Duggan, M. (2015). *Health online 2013*. 2013. URL: <http://pewinternet.org/Reports/2013/Health-online.aspx> [accessed 2015-04-25][WebCite Cache ID 6TymPhlbT].
- Fox, S. (2011). *The Social Life of Health Information, 2011*. Pew Internet & American Life Project, 1–33. <https://doi.org/http://www.who.int/topics/tuberculosis/en/>
- Frandsen, M., Walters, J., & Ferguson, S. G. (2014). Exploring the viability of using online social media advertising as a recruitment method for smoking cessation clinical trials. *Nicotine and Tobacco Research*, 16(2), 247–251. <https://doi.org/10.1093/ntr/ntt157>
- Freitas, C., Benevenuto, F., Ghosh, S., & Veloso, A. (2015, August). Reverse engineering socialbot infiltration strategies in twitter. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015* (pp. 25-32). ACM.
- Freyne, J., Berkovsky, S., Kimani, S., Baghaei, N., & Brindal, E. (2010). Improving health information access through social networking. In *International Symposium on Computer-Based Medical Systems (CBMS) IEEE (23rd : 12-15 oct. 2010: Perth, WA)* (pp. 334–339). <https://doi.org/10.1109/CBMS.2010.6042666>

- Friedman, D. B., Koskan, A., & Rose, I. D. (2011). Prostate cancer guidelines on Web 2.0-based sites: The screening dilemma continues online. In *Journal of Cancer Education* (Vol. 26, pp. 188–193). <https://doi.org/10.1007/s13187-010-0180-0>
- Gabarron, E., Serrano, J. A., Wynn, R., & Lau, A. Y. S. (2014). Tweet content related to sexually transmitted diseases: No joking matter. *Journal of Medical Internet Research*, 16(10), e228. <https://doi.org/10.2196/jmir.3259>
- Gardos, R. D., Burstein, A. D., Murtza, S., Hsian-Yeh, C., & Vajjala, A. (2015). U.S. Patent No. 9,106,712. Washington, DC: U.S. Patent and Trademark Office.
- George, A., Riddell, D., Seal, S., Talukdar, S., Mahamdallie, S., Ruark, E., ... Rahman, N. (2016). Implementing rapid, robust, cost-effective, patient-centred, routine genetic testing in ovarian cancer patients. *Scientific Reports*, 6(6).
- Gibbons, M. C., Fleisher, L., Slamon, R. E., Bass, S., Kandadai, V., & Beck, J. R. (2011). Exploring the potential of Web 2.0 to address health disparities. *Journal of health communication*, 16(sup1), 77-89.
- Giglietto, F., Rossi, L., & Bennato, D. (2012). The open laboratory: Limits and possibilities of using Facebook, Twitter, and YouTube as a research data source. *Journal of Technology in Human Services*, 30, 145–159.
- Glaser, B. G., Strauss, A. L., & Strutzel, E. (1968). The discovery of grounded theory; strategies for qualitative research. *Nursing research*, 17(4), 364.
- Google. (n.d.). How Search algorithms work. Retrieved November 14, 2018, from <https://www.google.com/search/howsearchworks/algorithms/>
- Gorman, K. S., Smith, A. M., Cimini, M. E., Halloran, K. M., & Lubiner, A. G. (2013). Reaching the Hard to Reach: Lessons Learned from a Statewide Outreach Initiative. *Journal of Community Practice*, 21(1–2), 105–123. <https://doi.org/10.1080/10705422.2013.788369>
- Graham, A. L., Milner, P., Saul, J. E., & Pfaff, L. (2008). Online advertising as a public health and recruitment tool: comparison of different media campaigns to increase demand for smoking cessation interventions.[Erratum appears in *J Med Internet Res*. 2009;11(1):e2]. *J Med Internet Res*, 10(5), e50. <https://doi.org/http://dx.doi.org/10.2196/jmir.1001>
- Greene, J. A., Choudhry, N. K., Kilabuk, E., & Shrank, W. H. (2011). Online social networking by patients with diabetes: A qualitative evaluation of communication with Facebook. *Journal of General Internal Medicine*, 26(3), 287–292. <https://doi.org/10.1007/s11606-010-1526-3>
- Greenwood, S., Perrin, A., & Duggan, M. (2016). Social media update 2016. *Pew Research Center*, 11, 83.
- Grewal, P., & Alagaratnam, S. (2013). The quality and readability of colorectal cancer information on the internet. *International Journal of Surgery*, 11(5), 410-413.
- Greysen, S. R., Kind, T., & Chretien, K. C. (2010). Online professionalism and the mirror of social media. *Journal of General Internal Medicine*. <https://doi.org/10.1007/s11606-010-1447-1>
- Griffiths, F., Cave, J., Boardman, F., Ren, J., Pawlikowska, T., Ball, R., ... Cohen, A. (2012). Social networks - The future for health care delivery. *Social Science & Medicine* (1982), 75(12), 2233–2241. <https://doi.org/10.1016/j.socscimed.2012.08.023>
- Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field methods*, 18(1), 59-82.

- Habibi, M. F., Nicklas, J., Spence, M., Hedberg, S., Magnuson, E., & Kavanagh, K. F. (2012). Remote Lactation Consultation: A Qualitative Study of Maternal Response to Experience and Recommendations for Survey Development. *Journal of Human Lactation*, 28(X), 211–217. <https://doi.org/10.1177/0890334411432716>
- Hagan, A., Gaither, C. A., & Bagozzi, R. P. (2010). The influence of support groups and identity on goal setting and achievement in persons with type 2 diabetes. *ProQuest Dissertations and Theses*.
- Heavilin, N., Gerbert, B., Page, J.E., & Gibbs, J.L. (2011). Public health surveillance of dental pain via Twitter. *Journal of Dental Research*, 90, 1047–1051.
- Heron, K. E., & Smyth, J. M. (2010). Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. *British Journal of Health Psychology*, 15(Pt 1), 1–39. <https://doi.org/10.1348/135910709X466063>
- Hesse BW, N. DE. (2005). Trust and sources of health information: The impact of the internet and its implications for health care providers: findings from the first health information national trends survey. *Archives of Internal Medicine*, 165(22), 2618–2624. <https://doi.org/10.1001/archinte.165.22.2618>
- Hesse BW, N. DE. (2005). Trust and sources of health information: The impact of the internet and its implications for health care providers: findings from the first health information national trends survey. *Archives of Internal Medicine*, 165(22), 2618–2624. <https://doi.org/10.1001/archinte.165.22.2618>
- Holtzblatt, K., Wendell, J. B., & Wood, S. (2005). *Rapid Contextual Design: A How-to Guide to Key Techniques for User-Centered Design*. Ubiquity. <https://doi.org/10.1145/1066322.1066325>
- Hong, Y., Peña-Purcell, N. C., & Ory, M. G. (2012). Outcomes of online support and resources for cancer survivors: A systematic literature review. *Patient Education and Counseling*. <https://doi.org/10.1016/j.pec.2011.06.014>
- Househ, M., Grainger, R., Petersen, C., Bamidis, P., & Merolli, M. (2018). Balancing Between Privacy and Patient Needs for Health Information in the Age of Participatory Health and Social Media: A Scoping Review. *Yearbook of medical informatics*.
- Howard, P. N. (2006). *New media campaigns and the managed citizen*. Cambridge University Press.
- Howlander, N., Noone, A. M., Krapcho, M., Miller, D., Bishop, K., Kosary, C. L., ... & Lewis, D. R. (2017). *SEER Cancer Statistics Review, 1975–2014*. Bethesda, MD: National Cancer Institute.
- Høybye, M. T., Johansen, C., & Tjørnhøj-Thomsen, T. (2005). Online interaction. Effects of storytelling in an internet breast cancer support group. *Psycho-Oncology*, 14(3), 211–220. <https://doi.org/10.1002/pon.837>
- Huerta, T., Walker, D., Johnson, T., & Ford, E. (2016). A Time Series Analysis of Cancer-Related Information Seeking: Hints From the Health Information National Trends Survey (HINTS) 2003–2014. *Journal of Health Communication*, 21(9), 1031–1038.
- Hughes, B., Joshi, I., Lemonde, H., & Wareham, J. (2009). Junior physician's use of Web 2.0 for information seeking and medical education: A qualitative study. *International Journal of Medical Informatics*, 78(10), 645–655. <https://doi.org/10.1016/j.ijmedinf.2009.04.008>
- Hwang, T., Pearce, I., & Nanis, M. (2012). Socialbots: Voices from the fronts. *interactions*, 19(2), 38–45.

- Im, E. O. (2011, August). Online support of patients and survivors of cancer. In *Seminars in oncology nursing* (Vol. 27, No. 3, pp. 229-236). WB Saunders.
- Impicciatore, P., Pandolfini, C., Casella, N., & Bonati, M. (1997). Reliability of health information for the public on the World Wide Web: systematic survey of advice on managing fever in children at home [see comments]. *British Medical Journal*, 314(7098), 1875–1879.
- Instagram (2018). <https://www.slideshare.net/wearesocial/digital-in-2018-global-overview-86860338>
- Intille, S. S., Kukla, C., Farzanfar, R., & Bakr, W. (2003). Just-in-time technology to encourage incremental, dietary behavior change. *AMIA ... Annual Symposium Proceedings / AMIA Symposium*. *AMIA Symposium*, 42(5), 874. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1480227&tool=pmcentrez&rendertype=abstract>
- Jennings, A., Powell, J., Armstrong, N., Sturt, J., & Dale, J. (2009). A virtual clinic for diabetes self-management: pilot study. *Journal of Medical Internet Research*, 11(1). <https://doi.org/10.2196/jmir.1111>
- Jin, D. Y., & Yoon, K. (2016). Reimagining smartphones in a local mediascape: A cultural analysis of young KakaoTalk users in Korea. *Convergence*, 22(5), 510-523.
- Johnson, J. D., & Meischke, H. (1993)A. Cancer-related channel selection: an extension for a sample of women who have had a mammogram. *Women & Health*, 20(2), 31–44. https://doi.org/10.1300/J013v20n02_03 [doi]
- Johnson, J. D., & Meischke, H. (1993)B. A Comprehensive Model of Cancer-Related Information Seeking Applied to Magazines. *Human Communication Research*, 19(3), 343–367. <https://doi.org/10.1111/j.1468-2958.1993.tb00305.x>
- Johnson, J. D., Donohue, W. A., Atkin, C. K., & Johnson, S. H. (1995). A comprehensive model of information seeking: Tests focusing on a technical organization. *Science Communication*, (16), 274–303.
- Johnson, J. D. (1997). *Cancer-related information seeking*. Cresskill, NJ: Hampton Press.
- Kantner, L., Sova, D. H., & Rosenbaum, S. (2003). Alternative methods for field usability research. *Proceedings of the 21st Annual International Conference on Documentation SIGDOC 03*Kantner, L., Sova, D. H., and Rosenbaum, S. 2003. “Alternative Methods for Field Usability Research,” *Proceedings of the 21st Annual International Conference on Documentatio*, 68. <https://doi.org/10.1145/944868.944883>
- Kaphingst, K. a, Lachance, C. R., & Condit, C. M. (2009). Beliefs about heritability of cancer and health information seeking and preventive behaviors. *Journal of Cancer Education : The Official Journal of the American Association for Cancer Education*, 24(4), 351–356. <https://doi.org/10.1080/08858190902876304>
- Kapp, J. M., Peters, C., & Oliver, D. P. (2013). Research recruitment using facebook advertising: Big potential, big challenges. *Journal of Cancer Education*, 28(1), 134–137. <https://doi.org/10.1007/s13187-012-0443-z>
- Keim, M. E., & Noji, E. (2011). Emergent use of social media: a new age of opportunity for disaster resilience. *American Journal of Disaster Medicine*, 6(1), 47–54. <https://doi.org/10.5055/ajdm.2010.0000>
- Keller, B., Labrique, A., Jain, K. M., Pekosz, A., & Levine, O. (2014). Mind the gap: Social media engagement by public health researchers. *Journal of Medical Internet Research*, 16(1). <https://doi.org/10.2196/jmir.2982>

- Kernisan, L. P., Sudore, R. L., & Knight, S. J. (2010). Information-seeking at a caregiving website: A qualitative analysis. *Journal of Medical Internet Research*, 12(3). <https://doi.org/10.2196/jmir.1548>
- Khoo, C. S. G. (2014). Issues in Information Behaviour on Social Media. *Proceedings of the ISIC Workshop on Information Behaviour on Social Media*, 24(2), 75–96.
- Kim, Y., Huang, J., & Emery, S. (2016). Garbage in, Garbage Out: Data Collection, Quality Assessment and Reporting Standards for Social Media Data Use in Health Research, Infodemiology and Digital Disease Detection. *Journal of Medical Internet Research*, 18(2), e41. <https://doi.org/10.2196/jmir.4738>
- Kim, S. (2009). Content analysis of cancer blog posts. *Journal of the Medical Library Association : JMLA*, 97(4), 260–6. <https://doi.org/10.3163/1536-5050.97.4.009>
- King, D., Ramirez-Cano, D., Greaves, F., Vlaev, I., Beales, S., & Darzi, A. (2013). Twitter and the health reforms in the English National Health Service. *Health Policy*, 110, 291–297.
- Konecki, K. T. (2011). Visual Grounded Theory: A Methodological Outline and Examples from Empirical Work. *Revija Za Sociologiju*, 41(2), 131–160. <https://doi.org/10.5613/rzs.41.2>.
- Kopits, I. M., Chen, C., Roberts, J. S., Uhlmann, W., & Green, R. C. (2011). Willingness to pay for genetic testing for Alzheimer’s disease: a measure of personal utility. *Genetic Testing and Molecular Biomarkers*, 15(12), 871–5. <https://doi.org/10.1089/gtmb.2011.0028>
- Krogstad, J. M. (2015). Social media preferences vary by race and ethnicity. Pew Research Center.
- Kuchenbaecker, K. B., Hopper, J. L., Barnes, D. R., Phillips, K. A., Mooij, T. M., Roos-Blom, M. J., ... & Goldgar, D. E. (2017). Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers. *Jama*, 317(23), 2402-2416.
- Kuehn, B. M. (2013). More than one-third of US individuals use the Internet to self-diagnose. *JAMA : The Journal of the American Medical Association*, 309, 756–7. <https://doi.org/10.1001/jama.2013.629>
- Kukreja, P., Heck Sheehan, A., & Riggins, J. (2011). Use of social media by pharmacy preceptors. *American Journal of Pharmaceutical Education*, 75(9), 176. <https://doi.org/10.5688/ajpe759176>
- Kushniruk, A. W., & Patel, V. L. (2004). Cognitive and usability engineering methods for the evaluation of clinical information systems. *Journal of Biomedical Informatics*, 37(1), 56–76. <https://doi.org/10.1016/j.jbi.2004.01.003> [doi]
- Lafferty, N. T., & Manca, A. (2018). Perspectives on social media in and as research: A synthetic review. In *Social Media in Medicine* (pp. 3-14). Routledge.
- Lambert, K. M., Barry, P., & Stokes, G. (2012). Risk management and legal issues with the use of social media in the healthcare setting. *J Healthc Risk Manag*, 31, 41–47. <https://doi.org/10.1002/jhrm.20103>
- Larson, H. J., Cooper, L. Z., Eskola, J., Katz, S. L., & Ratzan, S. (2011). Addressing the vaccine confidence gap. *The Lancet*, 378(9790), 526–535. [https://doi.org/10.1016/S0140-6736\(11\)60678-8](https://doi.org/10.1016/S0140-6736(11)60678-8)
- Lau, A. Y. S., Gabarron, E., Fernandez-Luque, L., & Armayones, M. (2012). Social media in health--what are the safety concerns for health consumers? *The HIM Journal*, 41(2), 30–5. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/23705132>
- Lawrentschuk, N., Sasges, D., Tasevski, R., Abouassaly, R., Scott, A. M., & Davis, I. D. (2012). Oncology health information quality on the Internet: a multilingual evaluation. *Annals of surgical oncology*, 19(3), 706-713.

- Lee, C. J., Gray, S. W., & Lewis, N. (2010). Internet use leads cancer patients to be active health care consumers. *Patient Education and Counseling*, 81(SUPPL. 1), S63–S69. <https://doi.org/10.1016/j.pec.2010.09.004>
- Lee, I. (2015). Health-seeking behavior of Korean women with myocardial infarction. In 48th Annual Communicating Nursing Research Conference.
- Leininger, M. (1994). Evaluation criteria and critique of qualitative studies. In *Critical issues in qualitative research methods* (pp. 95–115).
- Lemire, M., Sicotte, C., & Paré, G. (2008). Internet use and the logics of personal empowerment in health. *Health Policy*, 88(1), 130–140. <https://doi.org/10.1016/j.healthpol.2008.03.006>
- Lewis, S. C., Zamith, R., & Hermida, A. (2013). Content Analysis in an Era of Big Data: A Hybrid Approach to Computational and Manual Methods. *Journal of Broadcasting & Electronic Media*, 57(1), 34–52. <https://doi.org/10.1080/08838151.2012.761702>
- Luker, K. A., Beaver, K., Leinster, S. J., Owens, R. G., Degner, L. F., & Sloan, J. A. (1995). The information needs of women newly diagnosed with breast cancer. *Journal of Advanced Nursing*, 22(1), 134–141.
- Lyles, C., Lopez, A., Pasick, R., & Sarkar, U. (2013). “5 Mins of Uncomfyness Is Better than Dealing with Cancer 4 a Lifetime”: an Exploratory Qualitative Analysis of Cervical and Breast Cancer Screening Dialogue on Twitter. *Journal Of Cancer Education*, 28(1), 127–133.
- MacMillan, C. (2012). Social media revolution and blurring of professional boundaries. *Imprint*, 60(3), 44–46.
- Macnab, I., Rojjanasrirat, W., & Sanders, A. (2012). Breastfeeding and telehealth. *Journal of Human Lactation*, 28(4), 446–449.
- Maddock, C., Lewis, I., Ahmad, K., & Sullivan, R. (2011). Online information needs of cancer patients and their organizations. *Ecancermedicalsecience*, 5.
- Makri, S., Blandford, A., Gow, J., Rimmer, J., Warwick, C., & Buchanan, G. (2007). A library or just another information resource? A case study of users’ mental models of traditional and digital libraries. *Journal of the American Society for Information Science and Technology*, 58(3), 433–445.
- Martin-Facklam, M., Kostrzewa, M., Schubert, F., Gasse, C., & Haefeli, W. E. (2002). Quality markers of drug information on the Internet: an evaluation of sites about St. John’s wort. *The American Journal of Medicine*, 113(9), 740–745.
- Martinez, O., Shultz, A. Z., Rios, J. L., Sandfort, T., Carballo-Diequez, A., Remien, R. H., ... Rhodes, S. D. (2014). Still a hard-to-reach population? Using social media to recruit latino gay couples for an HIV intervention adaptation study. *Journal of Medical Internet Research*, 16(4).
- Mast, C., Wang, F., Su, S., Seeger, J., Kinyanjui, T. M., House, T. A., ... Hutson, S. (2015). Herd Immunity : Does Social Media Affect Adherence to the CDC Childhood Vaccination Schedule ? Pursuit-The Journal of Undergraduate Research at the University of Tennessee, 6.1(5), 9–21. <https://doi.org/10.1371/journal.pone.0138018>
- McCarney, R., Warner, J., Iliffe, S., Van Haselen, R., Griffin, M., & Fisher, P. (2007). The Hawthorne Effect: a randomised, controlled trial. *BMC medical research methodology*, 7(1), 30.
- McCormick, T. H., Lee, H., Cesare, N., Shojaie, A., & Spiro, E. S. (2015). Using Twitter for Demographic and Social Science Research Tools for Data Collection and Processing. *Sociological Methods & Research*.

- McIntosh, S., Ossip-Klein, D. J., Spada, J., & Burton, K. (2000). Recruitment strategies and success in a multi-county smoking cessation study. *Nicotine & Tobacco Research*, 2(3), 281–284.
- McKnight, M. (2014). Information prescriptions, 1930–2013: an international history and comprehensive review. *Journal of the Medical Library Association: JMLA*, 102(4), 271.
- McLaughlin, M., Nam, Y., Gould, J., Pade, C., Meeske, K. A., Ruccione, K. S., & Fulk, J. (2012). A videosharing social networking intervention for young adult cancer survivors. *Computers in Human Behavior*, 28(2), 631–641.
<https://doi.org/10.1016/j.chb.2011.11.009>
- Merchant, R. M., Elmer, S., & Lurie, N. (2011). Integrating social media into emergency-preparedness efforts. *New England Journal of Medicine*, 365(4), 289–291.
- Meric, F. (2002). Breast cancer on the world wide web: cross sectional survey of quality of information and popularity of websites. *Bmj*, 324(7337), 577–581.
<https://doi.org/10.1136/bmj.324.7337.577>
- MESM. (2017, February). Using social media for clinical trial recruitment ... Retrieved June 1, 2018, from <https://pharmaphorum.com/views-and-analysis/using-social-media-for-clinical-trial-recruitment/>
- Messias, J., Schmidt, L., Oliveira, R. A. R. D., & Benevenuto, F. R. (2013). You followed my bot! Transforming robots into influential users in Twitter.
- Metcalfe, A., Werrett, J., Burgess, L., Chapman, C., & Clifford, C. (2009). Cancer genetic predisposition: information needs of patients irrespective of risk level. *Familial Cancer*, 8(4), 403–412.
- Miller, L. M. S., & Bell, R. a. (2012). Online health information seeking: the influence of age, information trustworthiness, and search challenges. *Journal of Aging and Health*, 24(3), 525–41. <https://doi.org/10.1177/0898264311428167>
- Mills, R., Powell, J., Barry, W., & Haga, S. B. (2015). Information-Seeking and Sharing Behavior Following Genomic Testing for Diabetes Risk. *Journal of Genetic Counseling*, 24(1), 58–66.
- Mitchell, A. (2014). Which news organization is the most trusted? The answer is complicated. Pew Research Center.
- Mitra, T., Counts, S., & Pennebaker, J. W. (2016). Understanding Anti-Vaccination Attitudes in Social Media. In Tenth International AAAI Conference on Web and Social Media.
- Moen, A., Smørðal, O., & Sem, I. (2009). Web-based resources for peer support - Opportunities and challenges. In *Studies in Health Technology and Informatics* (Vol. 150, pp. 302–306). <https://doi.org/10.3233/978-1-60750-044-5-302>
- Monnier, J., Laken, M., & Carter, C. L. (2002). Patient and caregiver interest in internet-based cancer services. *Cancer Practice*, 10(6), 305–310.
- Moore, R. A., Matthews, A. L., & Cohen, L. (2018). How might the genetics profession better utilize social media. *Journal of genetic counseling*, 27(2), 493–504.
- Moorhead, S. A., Hazlett, D. E., Harrison, L., Carroll, J. K., Irwin, A., & Hoving, C. (2013). A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. *J Med Internet Res*, 15(4), e85.
<https://doi.org/10.2196/jmir.1933>
- Moorhead, S. A., Hazlett, D. E., Harrison, L., Carroll, J. K., Irwin, A., & Hoving, C. (2013). A new dimension of health care: systematic review of the uses, benefits, and limitations of

- social media for health communication. *J Med Internet Res*, 15(4), e85.
<https://doi.org/10.2196/jmir.1933>
- Morgan, T., Schmidt, J., Haakonsen, C., Lewis, J., Rocca, M. Della, Morrison, S., ... Kaphingst, K. A. (2014). Using the internet to seek information about genetic and rare diseases: A case study comparing data from 2006 and 2011. *Journal of Medical Internet Research*, 16(2). <https://doi.org/10.2196/resprot.2916>
- Mostaghimi, A., & Crotty, B. H. (2011). Professionalism in the digital age. *Annals of Internal Medicine*, 154(8), 560–562. <https://doi.org/10.1059/0003-4819-154-8-201104190-00008>
- Moyer, V. A. (2014). Risk assessment, genetic counseling, and genetic testing for BRCA-related cancer in women: US Preventive Services Task Force recommendation statement. *Annals of internal medicine*, 160(4), 271-281.
- Murray, E., Lo, B., Pollack, L., Donelan, K., Catania, J., White, M., ... Turner, R. (2003). The impact of health information on the internet on the physician-patient relationship: patient perceptions. *Arch Intern Med*, 163(14), 1727–1734.
<https://doi.org/10.1001/archinte.163.14.1727>
- Musiat, P., Winsall, M., Orlowski, S., Antezana, G., Geoffrey, S., Battersby, M., & Bidargaddi, N. (2016). Paid and Unpaid Online Recruitment for Health Interventions in Young Adults. *Journal of Adolescent Health*.
- National Cancer Institute. (2018) BRCA Mutations: Cancer Risk & Genetic Testing. (n.d.). Retrieved June 6, 2018, from <https://www.cancer.gov/about-cancer/causes-prevention/genetics/brca-fact-sheet>
- National Cancer Institute. (2018). BRCA Mutations: Cancer Risk & Genetic Testing. Retrieved from <https://www.cancer.gov/about-cancer/causes-prevention/genetics/brca-fact-sheet#r17>
- National Center for Advancing Translational Sciences (NCATS). 2018. Trial Innovation Network URL: <https://ncats.nih.gov/ctsa/projects/network>
- National Center for Advancing Translational Sciences (NCATS). 2018. CTSA consortium tackling clinical trial recruitment roadblocks URL: <https://ncats.nih.gov/pubs/features/ctsa-act>
- National Institutes of Health. (2011). How To Evaluate Health Information on the Internet: Questions and Answers. Retrieved July 4, 2018, from https://ods.od.nih.gov/Health_Information/How_To_Evaluate_Health_Information_on_the_Internet_Questions_and_Answers.aspx
- Nelissen, S., Beullens, K., Lemal, M., & Van den Bulck, J. (2015). Fear of cancer is associated with cancer information seeking, scanning and avoiding: a cross-sectional study among cancer diagnosed and non-diagnosed individuals. *Health Information & Libraries Journal*, 32(2), 107-119.
- Newby, R., Brodribb, W., Ware, R. S., & Davies, P. S. W. (2015). Internet Use by First-Time Mothers for Infant Feeding Support. *Journal of Human Lactation : Official Journal of International Lactation Consultant Association*, 31(3), 416–24.
<https://doi.org/10.1177/0890334415584319>
- Newington, L. ., & Metcalfe, A. . (2014). Factors influencing recruitment to research: Qualitative study of the experiences and perceptions of research teams. *BMC Medical Research Methodology*, 14(1). <https://doi.org/10.1186/1471-2288-14-10>

- Niederdeppe, J., Hornik, R., Kelly, B., Frosch, D., Romantan, A., Stevens, R., ... Schwartz, J. (2007). Examining the dimensions of cancer-related information scanning and seeking behavior. *Health Communication, 22*(2), 153–1.
- Nielsen Holdings. (n.d.). Global Consumers More Likely to Buy New Products from Familiar Brands. Retrieved August 12, 2018, from <https://www.nielsen.com/us/en/press-room/2013/global-consumers-more-likely-to-buy-new-products-from-familiar-b0.html>
- Noh, H. I., Lee, J. M., Yun, Y. H., Park, S. Y., Bae, D. S., Nam, J. H., ... Chang, Y. J. (2009). Cervical cancer patient information-seeking behaviors, information needs, and information sources in South Korea. *Supportive Care in Cancer, 17*(10), 1277–1283. <https://doi.org/10.1007/s00520-009-0581-y>
- Nölke, L., Mensing, M., Krämer, A., & Hornberg, C. (2015). Sociodemographic and health- (care-) related characteristics of online health information seekers: a cross-sectional German study. *BMC public health, 15*(1), 31.
- Nordqvist, C., Hanberger, L., Timpka, T., & Nordfeldt, S. (2009). Health professionals' attitudes towards using a web 2.0 portal for child and adolescent diabetes care: Qualitative study. *Journal of Medical Internet Research, 11*(2). <https://doi.org/10.2196/jmir.1152>
- Norman, C. D., & Skinner, H. A. (2006). eHealth literacy: Essential skills for consumer health in a networked world. *Journal of Medical Internet Research*. <https://doi.org/10.2196/jmir.8.2.e9>
- Norquist, B. M., Pennington, K. P., Agnew, K. J., Harrell, M. I., Pennil, C. C., Lee, M. K., ... Swisher, E. M. (2013). Characteristics of women with ovarian carcinoma who have BRCA1 and BRCA2 mutations not identified by clinical testing. *Gynecologic Oncology, 128*(3), 483–487. <https://doi.org/10.1016/j.ygyno.2012.12.015>
- O'Connor, A., Jackson, L., Goldsmith, L., & Skirton, H. (2014). Can I get a retweet please? Health research recruitment and the Twittersphere. *Journal of advanced nursing, 70*(3), 599-609.
- Obar, J. A., & Wildman, S. S. (2015). Social media definition and the governance challenge-an introduction to the special issue.
- Oh, K. E., & Belkin, N. (2015). Personal Information Categorization: The Rigid, Fuzzy, and Flexible Categorizers. In *iConference 2015 Proceedings*.
- Okoniewski, A., Lee, E., Rodriguez, Y., Schnall, J., & Low, M. (2014). Health Information Seeking Behaviors of Ethnically Diverse Adolescents. *Journal of Immigrant and Minority Health, 16*(4), 652–660.
- Ordenes, F. V., Grewal, D., Ludwig, S., Ruyter, K. D., Mahr, D., Wetzels, M., & Kopalle, P. (2018). Cutting through content clutter: How speech and image acts drive consumer sharing of social media brand messages. *Journal of Consumer Research*.
- Orizio, G., Schulz, P., Gasparotti, C., Caimi, L., & Gelatti, U. (2010). The World of e-Patients: A Content Analysis of Online Social Networks Focusing on Diseases. *Telemedicine and E-Health, 16*(10), 1060–1066. <https://doi.org/10.1089/tmj.2010.0085>
- Ovadia, S. (2009). Exploring the potential of Twitter as a research tool. *Behavioral and Social Sciences Librarian, 28*, 202–205.
- Pal, T., Permuth-Wey, J., Betts, J. A., Krischer, J. P., Fiorica, J., Arango, H., ... Sutphen, R. (2005). BRCA1 and BRCA2 mutations account for a large proportion of ovarian carcinoma cases. *Cancer, 104*(12), 2807–2816. <https://doi.org/10.1002/cncr.21536>

- Patrick, K., Intille, S. S., & Zabinski, M. F. (2005). An ecological framework for cancer communication: Implications for research. *Journal of Medical Internet Research*. <https://doi.org/10.2196/jmir.7.3.e23>
- Paul, M. J., & Dredze, M. (2011). You are what you Tweet: Analyzing Twitter for public health. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 265–272. <https://doi.org/10.1.1.224.9974>
- Pedersen, E. R., & Kurz, J. (2016). Using Facebook for health-related research study recruitment and program delivery. *Current Opinion in Psychology*. <https://doi.org/10.1016/j.copsyc.2015.09.011>
- Penchansky, R., & Thomas, J. W. (1981). The concept of access: definition and relationship to consumer satisfaction. *Medical Care*, 19(2), 127–140. <https://doi.org/10.2307/3764310>
- Perrin, A. (2015). Social media usage. *Pew research center*, 52-68.
- Petzel, S. V, Vogel, R. I., McNeil, J., Leininger, A., Argenta, P. A., & Geller, M. A. (2014). Improving Referral for Genetic Risk Assessment in Ovarian Cancer Using an Electronic Medical Record System. *International Journal of Gynecological Cancer*, 24(6), 1003–1009. <https://doi.org/10.1097/IGC.0000000000000148>
- Pew Research Center. (2018). Social media fact sheet. Pew Research Center: Internet, Science & Tech. <http://www.pewinternet.org/fact-sheet/social-media/>
- Pew Research Center. (2018). Social Media Use in 2018. Pew Research Center: Internet, Science & Tech. <http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/>
- Pew Research Center. (2018). Mobile fact sheet. *Pew Research Center: Internet, Science & Tech*. <http://www.pewinternet.org/fact-sheet/mobile/>
- Powell, J., Inglis, N., Ronnie, J., & Large, S. (2011). The characteristics and motivations of online health information seekers: cross-sectional survey and qualitative interview study. *Journal of medical Internet research*, 13(1).
- Quattrociochi, W., Scala, A., & Sunstein, C. (2016). Echo Chambers of Facebook.
- Rabin, C., Horowitz, S., & Marcus, B. (2013). Recruiting young adult cancer survivors for behavioral research. *Journal of Clinical Psychology in Medical Settings*, 20(1), 33–36. <https://doi.org/10.1007/s10880-012-9317-0>
- Raven, M. E., & Flanders, A. (1996). Using contextual inquiry to learn about your audiences. *ACM SIGDOC Asterisk Journal of Computer Documentation*, 20(1), 1–13. <https://doi.org/10.1145/227614.227615>
- Rees, C. E., & Bath, P. a. (2001). Information-seeking behaviors of women with breast cancer. *Oncology Nursing Forum*, 28(5), 899–907.
- Reuter, K., Angyan, P., Le, N., MacLennan, A., Cole, S., Bluthenthal, R. N., ... & Buchanan, T. A. (2018). Monitoring Twitter Conversations for Targeted Recruitment in Cancer Trials in Los Angeles County: Protocol for a Mixed-Methods Pilot Study. *JMIR research protocols*, 7(9), e177.
- Ricci, M. T., Sciallero, S., Mammoliti, S., Gismondi, V., Franiuk, M., Bruzzi, P., & Varesco, L. (2015). Referral of Ovarian Cancer Patients for Genetic Counselling by Oncologists: Need for Improvement. *Public Health Genomics*, 18(4), 225–232. <https://doi.org/10.1159/000431352>
- Richardson, J. T. (2011). The analysis of 2× 2 contingency tables— Yet again. *Statistics in medicine*, 30(8), 890-890.
- Roach, A. R., Lykins, E. L. B., Gochett, C. G., Brechting, E. H., Graue, L. O., & Andrykowski, M. A. (2009). Differences in cancer information seeking behavior, preferences, and

- awareness between cancer survivors and healthy controls: A national, population-based survey. *Journal of Cancer Education*, 24(1), 73–79.
<https://doi.org/10.1080/08858190802664784>.Differences
- Rocha, H. M., Savatt, J. M., Riggs, E. R., Wagner, J. K., Faucett, W. A., & Martin, C. L. (2018). Incorporating Social Media into your Support Tool Box: Points to Consider from Genetics-Based Communities. *Journal of genetic counseling*, 27(2), 470–480.
- Rojjanasrirat, W., Nelson, E.-L., & Wambach, K. a. (2012). A Pilot Study of Home-Based Videoconferencing for Breastfeeding Support. *Journal of Human Lactation*, 28(4), 464–467. <https://doi.org/10.1177/0890334412449071>
- Rony, M. M. U., Hassan, N., & Yousuf, M. (2017, July). Diving Deep into Clickbaits: Who Use Them to What Extents in Which Topics with What Effects?. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017* (pp. 232–239). ACM.
- Rosenstock, I. M. (1974). Historical origins of the health belief model. *Health Education & Behavior*, 2(4), 328–335. <https://doi.org/10.1177/109019817400200403>
- Rubin, A. M., & Windahl, S. (1986). The Uses and dependency model of mass communication. *Critical Studies in Mass Communication*, 3(2), 184–199.
<https://doi.org/10.1080/15295039609366643>
- Rus, H. M., & Cameron, L. D. (2016). Health Communication in Social Media: Message Features Predicting User Engagement on Diabetes-Related Facebook Pages. *Annals of Behavioral Medicine*, pp. 1–12. <https://doi.org/10.1007/s12160-016-9793-9>
- Rutten, L. J., Arora, N. K., Bakos, A. D., Aziz, N., & Rowland, J. (2005). Information needs and sources of information among cancer patients: a systematic review of research (1980–2003). *Patient Educ Couns*, 57(3), 250–261. <https://doi.org/10.1016/j.pec.2004.06.006>
- Rutten, L. J. F., Agunwamba, A. A., Wilson, P., Chawla, N., Vieux, S., Blanch-Hartigan, D., ... & Hesse, B. W. (2016). Cancer-related information seeking among cancer survivors: trends over a decade (2003–2013). *Journal of Cancer Education*, 31(2), 348–357.
- Sadler, G. R., Lee, H. C., Lim, R. S. H., & Fullerton, J. (2010). Recruitment of hard-to-reach population subgroups via adaptations of the snowball sampling strategy. *Nursing and Health Sciences*, 12(3), 369–374. <https://doi.org/10.1111/j.1442-2018.2010.00541.x>
- Sassenberg, K., & Greving, H. (2016). Internet Searching About Disease Elicits a Positive Perception of Own Health When Severity of Illness Is High: A Longitudinal Questionnaire Study. *Journal of Medical Internet Research*, 18(3), e56.
<https://doi.org/10.2196/jmir.5140>
- Saulsberry, K., & Terry, S. F. (2013). The need to build trust: A perspective on disparities in genetic testing. *Genetic testing and molecular biomarkers*, 17(9), 647–648.
- Scanfeld, D., Scanfeld, V., & Larson, E. L. (2010). Dissemination of health information through social networks: Twitter and antibiotics. *American Journal of Infection Control*, 38(3), 182–188. <https://doi.org/10.1016/j.ajic.2009.11.004>
- Schmutz, S., Sonderegger, A., & Sauer, J. (2016). Implementing recommendations from Web Accessibility Guidelines: Would they also provide benefits to nondisabled users. *Human factors*, 58(4), 611–629.
- Schwartz, M. D., Valdimarsdottir, H. B., Peshkin, B. N., Mandelblatt, J., Nusbaum, R., Huang, A. T., ... King, L. (2014). Randomized noninferiority trial of telephone versus in-person genetic counseling for hereditary breast and ovarian cancer. *Journal of Clinical Oncology*, 32(7), 618–626. <https://doi.org/10.1200/JCO.2013.51.3226>

- Seltzer, E. D., Stolley, M. R., Mensah, E. K., & Sharp, L. K. (2014). Social Networking Site Usage among Childhood Cancer Survivors-A Potential Tool for Research Recruitment? *Journal of Cancer Survivorship*, 8(3), 349–354. <https://doi.org/10.1007/s11764-014-0348-4>
- Setoyama, Y., Yamazaki, Y., & Namayama, K. (2011). Benefits of peer support in online Japanese breast cancer communities: differences between lurkers and posters. *Journal of Medical Internet Research*, 13(4). <https://doi.org/10.2196/jmir.1696>
- Shao, C., Ciampaglia, G. L., Varol, O., Yang, K., Flammini, A., & Menczer, F. (2017). The spread of low-credibility content by social bots. arXiv preprint arXiv:1707.07592.
- Shen, M. J., Dyson, R. C., D'agostino, T. A., Ostroff, J. S., Dickler, M. N., Heerdt, A. S., & Bylund, C. L. (2015). Cancer-related internet information communication between oncologists and patients with breast cancer: a qualitative study. *Psycho-Oncology*, 24(11), 1439-1447.
- Shigaki, C. L., Smarr, K. L., Gong, Y., Donovan-Hanson, K., Siva, C., Johnson, R. a, ... Musser, D. R. (2008). Social interactions in an online self-management program for rheumatoid arthritis. *Chronic Illness*, 4(4), 239–46. <https://doi.org/10.1177/1742395308097862>
- Siliquini, R., Ceruti, M., Lovato, E., Bert, F., Bruno, S., De Vito, E., ... La Torre, G. (2011). Surfing the internet for health information: an italian survey on use and population choices. *BMC Medical Informatics and Decision Making*, 11(1), 21. <https://doi.org/10.1186/1472-6947-11-21>
- Silva, B. M., Rodrigues, J. J., de la Torre Díez, I., López-Coronado, M., & Saleem, K. (2015). Mobile-health: A review of current state in 2015. *Journal of biomedical informatics*, 56, 265-272.
- Singh, A., Singh, S., & Singh, P. (2012). YouTube for Information on Rheumatoid Arthritis — A Wakeup Call? *The Journal of Rheumatology*, 39(5), 899–903.
- Slade, I., Riddell, D., Turnbull, C., Hanson, H., & Rahman, N. (2015). Development of cancer genetic services in the UK: A national consultation. *Genome Medicine*, 7(1), 18. <https://doi.org/10.1186/s13073-015-0128-4>
- Smith, A., & Brenner, J. (2012). Twitter use 2012. Pew Internet & American Life Project, 4.
- So, J., Prestin, A., Lee, L., Wang, Y., Yen, J., & Chou, W. Y. S. (2016). What do people like to “share” about obesity? A content analysis of frequent retweets about obesity on Twitter. *Health Communication*, 31(2), 193-206.
- Sobota, A., & Ozakinci, G. (2015). The quality and readability of online consumer information about gynecologic cancer. *Int J Gynecol Cancer*, 25(3), 537–541. <https://doi.org/10.1097/igc.0000000000000362>
- Soyano, A. E., Baldeo, C., & Kasi, P. M. (2018). BRCA Mutation and Its Association With Colorectal Cancer. *Clinical colorectal cancer*, 1.
- St-Pierre, D., Bouchard, K., Gauthier, L., Chiquette, J., & Dorval, M. (2018). Perspectives of Women Considering Bilateral Prophylactic Mastectomy and their Peers towards a Telephone-Based Peer Support Intervention. *Journal of genetic counseling*, 27(1), 274-288.
- Statista. (2018). United States: Number of Facebook users 2022 | Statistic. Retrieved September 14, 2018, from <https://www.statista.com/statistics/408971/number-of-us-facebook-users/>
- Steelman, T. A., McCaffrey, S. M., Velez, A. L. K., & Briefel, J. A. (2015). What information do people use, trust, and find useful during a disaster? Evidence from five large wildfires. *Natural Hazards*, 76(1), 615-634.

- Suh, B., Hong, L., Pirolli, P., & Chi, E. H. (2010, August). Want to be retweeted? large scale analytics on factors impacting retweet in twitter network. In 2010 IEEE Second International Conference on Social Computing (pp. 177-184). IEEE.
- Syed-Abdul, S., Fernandez-Luque, L., Jian, W. S., Li, Y. C., Crain, S., Hsu, M. H., ... Liou, D. M. (2013). Misleading health-related information promoted through video-based social media: Anorexia on youtube. *Journal of Medical Internet Research*, 15(2).
<https://doi.org/10.2196/jmir.2237>
- te Poel, F., Baumgartner, S. E., Hartmann, T., & Tanis, M. (2016). The curious case of cyberchondria: A longitudinal study on the reciprocal relationship between health anxiety and online health information seeking. *Journal of anxiety disorders*, 43, 32-40.
- Temple, E. C., & Brown, R. F. (2012). A comparison of internet-based participant recruitment methods: Engaging the hidden population of cannabis users in research. *Journal of Research Practice*, 7(2), 2.
- Tennant, B., Stellefson, M., Dodd, V., Chaney, B., Chaney, D., Paige, S., & Alber, J. (2015). eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. *Journal of medical Internet research*, 17(3).
- Terdiman, D. (2017, November 29). Facebook will no longer let users exclude racial groups in ad targeting. Retrieved from <https://www.fastcompany.com/40501775/facebook-will-no-longer-let-users-exclude-racial-groups-in-ad-targeting>
- The New York Times. (2018, April 10). Mark Zuckerberg Testimony: Senators Question Facebook's Commitment to Privacy. Retrieved from <https://www.nytimes.com/2018/04/10/us/politics/mark-zuckerberg-testimony.html>
- Theiss, S. K., Burke, R. M., Cory, J. L., & Fairley, T. L. (2016). Getting beyond impressions: an evaluation of engagement with breast cancer-related Facebook content. *MHealth*, 2.
- Tian, Y. (2010). Organ donation on Web 2.0: content and audience analysis of organ donation videos on YouTube. *Health Communication*, 25(3), 238-246.
<https://doi.org/10.1080/10410231003698911>
- Topolovec-Vranic, J., & Natarajan, K. (2016). The use of social media in recruitment for medical research studies: a scoping review. *Journal of medical Internet research*, 18(11).
- Treweek, S., Pitkethly, M., Cook, J., Fraser, C., Mitchell, E., Sullivan, F., ... & Gardner, H. (2018). Strategies to improve recruitment to randomised trials. *The Cochrane Library*.
- Tripp, C., Jensen, T. D., & Carlson, L. (1994). The effects of multiple product endorsements by celebrities on consumers' attitudes and intentions. *Journal of consumer research*, 20(4), 535-547.
- Trivers, K. F., Baldwin, L. M., Miller, J. W., Matthews, B., Andrilla, C. H. A., Lishner, D. M., & Goff, B. A. (2011). Reported referral for genetic counseling or BRCA 1/2 testing among United States physicians. *Cancer*, 117(23), 5334-5343.
- Tumblr . (2014) . Privacy policy . Retrieved 27 October 2014 from <https://www.tumblr.com/policy/en/privacy>
- Twitter . (2014b) . Twitter engineering blog –Introducing T witter data grants 05/02/2014 . Retrieved 27 October 2014 from <https://blog.twitter.com/2014/introducing-twitter-datagrants>.
- Twitter. (n.d.). Number of monthly active Twitter users in the United States from 1st quarter 2010 to 3rd quarter 2018 (in millions). Retrieved September 5, 2018, from <https://www.statista.com/statistics/274564/monthly-active-twitter-users-in-the-united-states/>

- Twitter. (n.d.). Help Center. Retrieved July 5, 2018, from <https://help.twitter.com/en>
- Twitter (2018). Selected Company Metrics and Financials. Retrieved November 2018. <https://investor.twitterinc.com/static-files/b9402133-be92-4ea4-ac2b-db20be19d1cd>
- Van Der Molen, B. (1999). Relating information needs to the cancer experience: 1. Information as a key coping strategy. *European Journal of Cancer Care*, 8(4), 238–244. <https://doi.org/10.1046/j.1365-2354.1999.00176.x>
- Varol, O., Ferrara, E., Menczer, F., & Flammini, A. (2017). Early detection of promoted campaigns on social media. *EPJ Data Science*, 6(1), 13.
- Varol, O., Ferrara, E., Davis, C. A., Menczer, F., & Flammini, A. (2017). Online human-bot interactions: Detection, estimation, and characterization. *arXiv preprint arXiv:1703.03107*.
- Ventola, C. L. (2014). Social media and health care professionals: benefits, risks, and best practices. *P & T : A Peer-Reviewed Journal for Formulary Management*, 39(7), 491–520. Retrieved from <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4103576&tool=pmcentrez&rendertype=abstract>
- Volkman, J. E., Luger, T. M., Harvey, K. L., Hogan, T. P., Shimada, S. L., Amante, D., ... & Houston, T. K. (2014). The National Cancer Institute's Health Information National Trends Survey [HINTS]: a national cross-sectional analysis of talking to your doctor and other healthcare providers for health information. *BMC family practice*, 15(1), 111.
- Voytek, C., Jones, K., Curtis, B., Fiore, D., Dunbar, D., & Frank, I. (2012). Using an Internet consumer marketing strategy to reach men who have sex with men for participation in a preventive HIV vaccine clinical trial. *Retrovirology*, 9, 119.
- Wable, A. (2010). Intro to facebook search. <https://www.facebook.com/notes/facebook-engineering/intro-to-facebook-search/365915113919>
- Walker, D. M. (2013). The internet as a medium for health services research. Part 2. *Nurse researcher*, 20(5).
- Walsh, T., Casadei, S., Coats, K. H., Swisher, E., Stray, S. M., Higgins, J., ... & Foretova, L. (2006). Spectrum of mutations in BRCA1, BRCA2, CHEK2, and TP53 in families at high risk of breast cancer. *Jama*, 295(12), 1379-1388.
- Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012, July). A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In *Proceedings of the ACL 2012 System Demonstrations* (pp. 115-120). Association for Computational Linguistics.
- Weaver, J. B., Mays, D., Weaver, S. S., Hopkins, G. L., Eroglu, D., & Bernhardt, J. M. (2010). Health information-seeking behaviors, health indicators, and health risks. *American Journal of Public Health*, 100(8), 1520–1525. <https://doi.org/10.2105/AJPH.2009.180521>
- Weinstein, N. D., & Klein, W. M. (1995). Resistance of Personal Risk Perceptions to Debiasing Interventions. *Health Psychology*, 14(2), 132–140. <https://doi.org/10.1037/0278-6133.14.2.132>
- Welch, B. M., O'Connell, N. S., Qanungo, S., Halbert-Hughes, C., & Schiffman, J. D. (2015). Collecting family health history using an online social network: A Nationwide survey among potential users. In *AMIA Annual Symposium Proceedings* (Vol. 2015, p. 1316). American Medical Informatics Association.

- Whitaker, C., Stevelink, S., & Fear, N. (2017). The use of Facebook in recruiting participants for health research purposes: a systematic review. *Journal of medical Internet research*, 19(8).
- Wojcik, S., Messing, S., Smith, A., Rainie, L., & Hitlin, P. (2018). Bots in the Twittersphere. Pew Research Center. Retrieved May, 22, 2018.
- Wong, P. W. ., Fu, K. ., Yau, R. S. ., Ma, H. H. ., Law, Y. ., Chang, S. ., & Yip, P. S. . (2013). Accessing suicide-related information on the internet: a retrospective observational study of search behavior. *Journal of Medical Internet Research*, 15(1), p.e3.
- Wong, N. C. (2012). Interaction of comparative cancer risk and cancer efficacy perceptions on cancer-related information seeking and scanning behaviors. *Communication Research Reports*, 29(3), 193–203.
- Yancey, A. K., Alexander, O. N., & Kumanyika, S. K. (2006). Effective recruitment and retention of minority research participants. *Annual Review Public Health*, 27, 1–28.
- Yin, Z., Fabbri, D., Rosenbloom, S. T., & Malin, B. (2015). A scalable framework to detect personal health mentions on Twitter. *Journal of Medical Internet Research*, 17(6), e138. <https://doi.org/10.2196/jmir.4305>
- Yuan, P., Bare, M. G., Johnson, M. O., & Saberi, P. (2014). Using online social media for recruitment of human immunodeficiency virus-positive participants: A cross-sectional survey. *Journal of Medical Internet Research*, 16(5). <https://doi.org/10.2196/jmir.3229>
- Zannettou, S., Sirivianos, M., Blackburn, J., & Kourtellis, N. (2018). The Web of False Information: Rumors, Fake News, Hoaxes, Clickbait, and Various Other Shenanigans. arXiv preprint arXiv:1804.03461.
- Zhang, S., & Cabage, N. (2017). Search engine optimization: Comparison of link building and social sharing. *Journal of Computer Information Systems*, 57(2), 148-159.
- Zhang, Y., Wang, H., Yin, G., Wang, T., & Yu, Y. (2017). Social media in GitHub: the role of@-mention in assisting software development. *Science China Information Sciences*, 60(3), 032102.
- Zulman, D. M., Kirch, M., Zheng, K., & An, L. C. (2011). Trust in the internet as a health resource among older adults: Analysis of data from a nationally representative survey. *Journal of Medical Internet Research*, 13(1). <https://doi.org/10.2196/jmir.1552>